

Exhibit A

Meeting January 14 1965

President's Address

The Environment and Disease: Association or Causation?

by Sir Austin Bradford Hill CBE DSC FRCP(hon) FRS
(*Professor Emeritus of Medical Statistics,
University of London*)

Amongst the objects of this newly-founded Section of Occupational Medicine are firstly 'to provide a means, not readily afforded elsewhere, whereby physicians and surgeons with a special knowledge of the relationship between sickness and injury and conditions of work may discuss their problems, not only with each other, but also with colleagues in other fields, by holding joint meetings with other Sections of the Society'; and, secondly, 'to make available information about the physical, chemical and psychological hazards of occupation, and in particular about those that are rare or not easily recognized'.

At this first meeting of the Section and before, with however laudable intentions, we set about instructing our colleagues in other fields, it will be proper to consider a problem fundamental to our own. How in the first place do we detect these relationships between sickness, injury and conditions of work? How do we determine what are physical, chemical and psychological hazards of occupation, and in particular those that are rare and not easily recognized?

There are, of course, instances in which we can reasonably answer these questions from the general body of medical knowledge. A particular, and perhaps extreme, physical environment cannot fail to be harmful; a particular chemical is known to be toxic to man and therefore suspect on the factory floor. Sometimes, alternatively, we may be able to consider what *might* a particular environment do to man, and then see whether such consequences are indeed to be found. But more often than not we have no such guidance, no such means of proceeding; more often than not we are dependent upon our observation and enumeration of defined events for which we then seek antecedents. In other words we see that the event B is associated with the environmental feature A, that, to take a specific example, some form of respiratory illness is associated with a dust in the environment. In what circumstances can we pass from this

observed *association* to a verdict of *causation*? Upon what basis should we proceed to do so?

I have no wish, nor the skill, to embark upon a philosophical discussion of the meaning of 'causation'. The 'cause' of illness may be immediate and direct, it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymously preventive, medicine in mind the decisive question is whether the frequency of the undesirable event B will be influenced by a change in the environmental feature A. How such a change exerts that influence may call for a great deal of research. However, before deducing 'causation' and taking action we shall not invariably have to sit around awaiting the results of that research. The whole chain may have to be unravelled or a few links may suffice. It will depend upon circumstances.

Disregarding then any such problem in semantics we have this situation. Our observations reveal an association between two variables, perfectly clear-cut and beyond what we would care to attribute to the play of chance. What aspects of that association should we especially consider before deciding that the most likely interpretation of it is causation?

(1) **Strength.** First upon my list I would put the strength of the association. To take a very old example, by comparing the occupations of patients with scrotal cancer with the occupations of patients presenting with other diseases, Percival Pott could reach a correct conclusion because of the *enormous* increase of scrotal cancer in the chimney sweeps. 'Even as late as the second decade of the twentieth century', writes Richard Doll (1964), 'the mortality of chimney sweeps from scrotal cancer was some 200 times that of workers who were not specially exposed to tar or mineral oils and in the eighteenth century the relative difference is likely to have been much greater.'

To take a more modern and more general example upon which I have now reflected for over fifteen years, prospective inquiries into smoking have shown that the death rate from cancer of the lung in cigarette smokers is nine to ten times the rate in non-smokers and the rate in heavy cigarette smokers is twenty to thirty times

as great. On the other hand the death rate from coronary thrombosis in smokers is no more than twice, possibly less, the death rate in non-smokers. Though there is good evidence to support causation it is surely much easier in this case to think of some features of life that may go hand-in-hand with smoking – features that might conceivably be the real underlying cause or, at the least, an important contributor, whether it be lack of exercise, nature of diet or other factors. But to explain the pronounced excess in cancer of the lung in any other environmental terms requires some feature of life so intimately linked with cigarette smoking and with the amount of smoking that such a feature should be easily detectable. If we cannot detect it or reasonably infer a specific one, then in such circumstances I think we are reasonably entitled to reject the vague contention of the armchair critic 'you can't prove it, there *may* be such a feature'.

Certainly in this situation I would reject the argument sometimes advanced that what matters is the absolute difference between the death rates of our various groups and not the ratio of one to other. That depends upon what we want to know. If we want to know how many extra deaths from cancer of the lung will take place through smoking (i.e. presuming causation), then obviously we must use the absolute differences between the death rates – 0.07 per 1,000 per year in non-smoking doctors, 0.57 in those smoking 1–14 cigarettes daily, 1.39 for 15–24 cigarettes daily and 2.27 for 25 or more daily. But it does not follow here, or in more specifically occupational problems, that this best measure of the effect upon mortality is also the best measure in relation to aetiology. In this respect the ratios of 8, 20 and 32 to 1 are far more informative. It does not, of course, follow that the differences revealed by ratios are of any practical importance. Maybe they are, maybe they are not; but that is another point altogether.

We may recall John Snow's classic analysis of the opening weeks of the cholera epidemic of 1854 (Snow 1855). The death rate that he recorded in the customers supplied with the grossly polluted water of the Southwark and Vauxhall Company was in truth quite low – 71 deaths in each 10,000 houses. What stands out vividly is the fact that the small rate is 14 times the figure of 5 deaths per 10,000 houses supplied with the sewage-free water of the rival Lambeth Company.

In thus putting emphasis upon the strength of an association we must, nevertheless, look at the obverse of the coin. We must not be too ready to dismiss a cause-and-effect hypothesis merely on

the grounds that the observed association appears to be slight. There are many occasions in medicine when this is in truth so. Relatively few persons harbouring the meningococcus fall sick of meningococcal meningitis. Relatively few persons occupationally exposed to rat's urine contract Weil's disease.

(2) **Consistency:** Next on my list of features to be specially considered I would place the *consistency* of the observed association. Has it been repeatedly observed by different persons, in different places, circumstances and times?

This requirement may be of special importance for those rare hazards singled out in the Section's terms of reference. With many alert minds at work in industry today many an environmental association may be thrown up. Some of them on the customary tests of statistical significance will appear to be unlikely to be due to chance. Nevertheless whether chance is the explanation or whether a true hazard has been revealed may sometimes be answered only by a repetition of the circumstances and the observations.

Returning to my more general example, the Advisory Committee to the Surgeon-General of the United States Public Health Service found the association of smoking with cancer of the lung in 29 retrospective and 7 prospective inquiries (US Department of Health, Education & Welfare 1964). The lesson here is that broadly the same answer has been reached in quite a wide variety of situations and techniques. In other words we can justifiably infer that the association is not due to some constant error or fallacy that permeates every inquiry. And we have indeed to be on our guard against that.

Take, for instance, an example given by Heady (1958). Patients admitted to hospital for operation for peptic ulcer are questioned about recent domestic anxieties or crises that may have precipitated the acute illness. As controls, patients admitted for operation for a simple hernia are similarly quizzed. But, as Heady points out, the two groups may not be *in pari materia*. If your wife ran off with the lodger last week you still have to take your perforated ulcer to hospital without delay. But with a hernia you might prefer to stay at home for a while – to mourn (or celebrate) the event. No number of exact repetitions would remove or necessarily reveal that fallacy.

We have, therefore, the somewhat paradoxical position that the different results of a different inquiry certainly cannot be held to refute the

original evidence; yet the same results from precisely the same form of inquiry will not invariably greatly strengthen the original evidence. I would myself put a good deal of weight upon similar results reached in quite different ways, e.g. prospectively and retrospectively.

Once again looking at the obverse of the coin there will be occasions when repetition is absent or impossible and yet we should not hesitate to draw conclusions. The experience of the nickel refiners of South Wales is an outstanding example. I quote from the Alfred Watson Memorial Lecture that I gave in 1962 to the Institute of Actuaries:

'The population at risk, workers and pensioners, numbered about one thousand. During the ten years 1929 to 1938, sixteen of them had died from cancer of the lung, eleven of them had died from cancer of the nasal sinuses. At the age specific death rates of England and Wales at that time, one might have anticipated one death from cancer of the lung (to compare with the 16), and a fraction of a death from cancer of the nose (to compare with the 11). In all other bodily sites cancer had appeared on the death certificate 11 times and one would have expected it to do so 10-11 times. There had been 67 deaths from all other causes of mortality and over the ten years' period 72 would have been expected at the national death rates. Finally division of the population at risk in relation to their jobs showed that the excess of cancer of the lung and nose had fallen wholly upon the workers employed in the chemical processes.'

'More recently my colleague, Dr Richard Doll, has brought this story a stage further. In the nine years 1948 to 1956 there had been, he found, 48 deaths from cancer of the lung and 13 deaths from cancer of the nose. He assessed the numbers expected at normal rates of mortality as, respectively 10 and 0.1.'

'In 1923, long before any special hazard had been recognized, certain changes in the refinery took place. No case of cancer of the nose has been observed in any man who first entered the works after that year, and in these men there has been no excess of cancer of the lung. In other words, the excess in both sites is uniquely a feature in men who entered the refinery in, roughly, the first 23 years of the present century.'

'No causal agent of these neoplasms has been identified. Until recently no animal experimentation had given any clue or any support to this wholly statistical evidence. Yet I wonder if any of us would hesitate to accept it as proof of a grave industrial hazard?' (Hill 1962).

In relation to my present discussion I know of no parallel investigation. We have (or certainly had) to make up our minds on a unique event; and there is no difficulty in doing so.

(3) Specificity: One reason, needless to say, is the specificity of the association, the third characteristic which invariably we must consider. If, as here, the association is limited to specific workers and to particular sites and types of disease and there is no association between the work and other modes of dying, then clearly that is a strong argument in favour of causation.

We must not, however, over-emphasize the importance of the characteristic. Even in my present example there is a cause and effect relationship with two different sites of cancer – the lung and the nose. Milk as a carrier of infection and, in that sense, the cause of disease can produce such a disparate galaxy as scarlet fever, diphtheria, tuberculosis, undulant fever, sore throat, dysentery and typhoid fever. Before the discovery of the underlying factor, the bacterial origin of disease, harm would have been done by pushing too firmly the need for specificity as a necessary feature before convicting the dairy.

Coming to modern times the prospective investigations of smoking and cancer of the lung have been criticized for not showing specificity – in other words the death rate of smokers is higher than the death rate of non-smokers from many causes of death (though in fact the results of Doll & Hill, 1964, do not show that). But here surely one must return to my first characteristic, the strength of the association. If other causes of death are raised 10, 20 or even 50% in smokers whereas cancer of the lung is raised 900-1,000% we have specificity – a specificity in the magnitude of the association.

We must also keep in mind that diseases may have more than one cause. It has always been possible to acquire a cancer of the scrotum without sweeping chimneys or taking to mule-spinning in Lancashire. One-to-one relationships are not frequent. Indeed I believe that multi-causation is generally more likely than single causation though possibly if we knew all the answers we might get back to a single factor.

In short, if specificity exists we may be able to draw conclusions without hesitation; if it is not apparent, we are not thereby necessarily left sitting irresolutely on the fence.

(4) Temporality: My fourth characteristic is the temporal relationship of the association – which is the cart and which the horse? This is a question which might be particularly relevant with diseases of slow development. Does a particular diet lead to disease or do the early stages of the disease lead to those peculiar dietetic habits? Does a

particular occupation or occupational environment promote infection by the tubercle bacillus or are the men and women who select that kind of work more liable to contract tuberculosis whatever the environment – or, indeed, have they already contracted it? This temporal problem may not arise often but it certainly needs to be remembered, particularly with selective factors at work in industry.

(5) Biological gradient: Fifthly, if the association is one which can reveal a biological gradient, or dose-response curve, then we should look most carefully for such evidence. For instance, the fact that the death rate from cancer of the lung rises linearly with the number of cigarettes smoked daily, adds a very great deal to the simpler evidence that cigarette smokers have a higher death rate than non-smokers. That comparison would be weakened, though not necessarily destroyed, if it depended upon, say, a much heavier death rate in light smokers and a lower rate in heavier smokers. We should then need to envisage some much more complex relationship to satisfy the cause-and-effect hypothesis. The clear dose-response curve admits of a simple explanation and obviously puts the case in a clearer light.

The same would clearly be true of an alleged dust hazard in industry. The dustier the environment the greater the incidence of disease we would expect to see. Often the difficulty is to secure some satisfactory quantitative measure of the environment which will permit us to explore this dose-response. But we should invariably seek it.

(6) Plausibility: It will be helpful if the causation we suspect is biologically plausible. But this is a feature I am convinced we cannot demand. What is biologically plausible depends upon the biological knowledge of the day.

To quote again from my Alfred Watson Memorial Lecture (Hill 1962), there was

‘... no biological knowledge to support (or to refute) Pott’s observation in the 18th century of the excess of cancer in chimney sweeps. It was lack of biological knowledge in the 19th that led a prize essayist writing on the value and the fallacy of statistics to conclude, amongst other “absurd” associations, that “it could be no more ridiculous for the stranger who passed the night in the steerage of an emigrant ship to ascribe the typhus, which he there contracted, to the vermin with which bodies of the sick might be infected”. And coming to nearer times, in the 20th century there was no biological knowledge to support the evidence against rubella.’

In short, the association we observe may be one new to science or medicine and we must not dismiss it too light-heartedly as just too odd. As Sherlock Holmes advised Dr Watson, ‘when you have eliminated the impossible, whatever remains, however improbable, must be the truth.’

(7) Coherence: On the other hand the cause-and-effect interpretation of our data should not seriously conflict with the generally known facts of the natural history and biology of the disease – in the expression of the Advisory Committee to the Surgeon-General it should have coherence.

Thus in the discussion of lung cancer the Committee finds its association with cigarette smoking coherent with the temporal rise that has taken place in the two variables over the last generation and with the sex difference in mortality – features that might well apply in an occupational problem. The known urban/rural ratio of lung cancer mortality does not detract from coherence, nor the restriction of the effect to the lung.

Personally, I regard as greatly contributing to coherence the histopathological evidence from the bronchial epithelium of smokers and the isolation from cigarette smoke of factors carcinogenic for the skin of laboratory animals. Nevertheless, while such laboratory evidence can enormously strengthen the hypothesis and, indeed, may determine the actual causative agent, the lack of such evidence cannot nullify the epidemiological observations in man. Arsenic can undoubtedly cause cancer of the skin in man but it has never been possible to demonstrate such an effect on any other animal. In a wider field John Snow’s epidemiological observations on the conveyance of cholera by the water from the Broad Street pump would have been put almost beyond dispute if Robert Koch had been then around to isolate the vibrio from the baby’s nappies, the well itself and the gentleman in delicate health from Brighton. Yet the fact that Koch’s work was to be awaited another thirty years did not really weaken the epidemiological case though it made it more difficult to establish against the criticisms of the day – both just and unjust.

(8) Experiment: Occasionally it is possible to appeal to experimental, or semi-experimental, evidence. For example, because of an observed association some preventive action is taken. Does it in fact prevent? The dust in the workshop is reduced, lubricating oils are changed, persons stop smoking cigarettes. Is the frequency of the associated events affected? Here the strongest

support for the causation hypothesis may be revealed.

(9) **Analogy:** In some circumstances it would be fair to judge by analogy. With the effects of thalidomide and rubella before us we would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy.

Here then are nine different viewpoints from all of which we should study association before we cry causation. **What I do not believe – and this has been suggested – is that we can usefully lay down some hard-and-fast rules of evidence that must be obeyed before we accept cause and effect.** None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?

Tests of Significance

No formal tests of significance can answer those questions. Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects. Beyond that they contribute nothing to the 'proof' of our hypothesis.

Nearly forty years ago, amongst the studies of occupational health that I made for the Industrial Health Research Board of the Medical Research Council was one that concerned the workers in the cotton-spinning mills of Lancashire (Hill 1930). The question that I had to answer, by the use of the National Health Insurance records of that time, was this: Do the workers in the cardroom of the spinning mill, who tend the machines that clean the raw cotton, have a sickness experience in any way different from that of other operatives in the same mills who are relatively unexposed to the dust and fibre that were features of the cardroom? The answer was an unqualified 'Yes'. From age 30 to age 60 the cardroom workers suffered over three times as much from respiratory causes of illness whereas from non-respiratory causes their experience was not different from that of the other workers. This pronounced difference with the respiratory causes was derived not from abnormally long periods of sickness but rather from an excessive number of repeated absences from work of the cardroom workers.

All this has rightly passed into the limbo of forgotten things. What interests me today is this: My results were set out for men and women separately and for half a dozen age groups in 36 tables. So there were plenty of sums. Yet I cannot find that anywhere I thought it necessary to use a test of significance. The evidence was so clear-cut, the differences between the groups were mainly so large, the contrast between respiratory and non-respiratory causes of illness so specific, that no formal tests could really contribute anything of value to the argument. So why use them?

Would we think or act that way today? I rather doubt it. Between the two world wars there was a strong case for emphasizing to the clinician and other research workers the importance of not overlooking the effects of the play of chance upon their data. Perhaps too often generalities were based upon two men and a laboratory dog while the treatment of choice was deduced from a difference between two bedfuls of patients and might easily have no true meaning. It was therefore a useful corrective for statisticians to stress, and to teach the need for, tests of significance merely to serve as guides to caution before drawing a conclusion, before inflating the particular to the general.

I wonder whether the pendulum has not swung too far – not only with the attentive pupils but even with the statisticians themselves. To decline to draw conclusions without standard errors can surely be just as silly? Fortunately I believe we have not yet gone so far as our friends in the USA where, I am told, some editors of journals will return an article because tests of significance have not been applied. Yet there are innumerable situations in which they are totally unnecessary – because the difference is grotesquely obvious, because it is negligible, or because, whether it be formally significant or not, it is too small to be of any practical importance. What is worse the glitter of the *t* table diverts attention from the inadequacies of the fare. Only a tithe, and an unknown tithe, of the factory personnel volunteer for some procedure or interview, 20% of patients treated in some particular way are lost to sight, 30% of a randomly-drawn sample are never contacted. The sample may, indeed, be akin to that of the man who, according to Swift, 'had a mind to sell his house and carried a piece of brick in his pocket, which he showed as a pattern to encourage purchasers'. The writer, the editor and the reader are unmoved. The magic formulæ are there.

Of course I exaggerate. Yet too often I suspect we waste a deal of time, we grasp the shadow and

lose the substance, we weaken our capacity to interpret data and to take reasonable decisions whatever the value of P. And far too often we deduce 'no difference' from 'no significant difference'. Like fire, the χ^2 test is an excellent servant and a bad master.

The Case for Action

Finally, in passing from association to causation I believe in 'real life' we shall have to consider what flows from that decision. On scientific grounds we should do no such thing. The evidence is there to be judged on its merits and the judgment (in that sense) should be utterly independent of what hangs upon it – or who hangs because of it. But in another and more practical sense we may surely ask what is involved in our decision. In occupational medicine our object is usually to take action. If this be operative cause and that be deleterious effect, then we shall wish to intervene to abolish or reduce death or disease.

While that is a commendable ambition it almost inevitably leads us to introduce differential standards before we convict. Thus on relatively slight evidence we might decide to restrict the use of a drug for early-morning sickness in pregnant women. If we are wrong in deducing causation from association no great harm will be done. The good lady and the pharmaceutical industry will doubtless survive.

On fair evidence we might take action on what appears to be an occupational hazard, e.g. we might change from a probably carcinogenic oil

to a non-carcinogenic oil in a limited environment and without too much injustice if we are wrong. But we should need very strong evidence before we made people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and sugar that they do like. In asking for very strong evidence I would, however, repeat emphatically that this does not imply crossing every 't', and swords with every critic, before we act.

All scientific work is incomplete – whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.

Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8.30 next day.

REFERENCES

- Doll R (1964) In: *Medical Surveys and Clinical Trials*. Ed. L J Witts. 2nd ed. London; p 333
- Doll R & Hill A B (1964) *Brit. med. J.* i, 1399, 1460
- Heady J A (1958) *Med. World, Lond.* 89, 305
- Hill A B (1930) *Sickness amongst Operatives in Lancashire Spinning Mills*. Industrial Health Research Board Report No. 59. HMSO, London (1962) *J. Inst. Actu.* 88, 178
- Snow J (1855) *On the Mode of Communication of Cholera*. 2nd ed. London (Reprinted 1936, New York)
- US Department of Health, Education & Welfare (1964) *Smoking and Health*. Public Health Service Publication No. 1103. Washington

Exhibit B

Foundations of Epidemiology

THIRD EDITION

Revised by
David E. Lilienfeld Paul D. Stolley

RESERVE

Foundations of Epidemiology

THIRD EDITION

Revised by

David E. Lilienfeld

M.D., M.S. Eng., M.P.H., F.A.C.E.

Senior Epidemiologist, EMMES Corp.

Paul D. Stolley

M.D., M.P.H., F.A.C.E., F.A.C.P.

Professor and Chairman,

Department of Epidemiology
and Preventive Medicine

University of Maryland School of Medicine

Original edition by Abraham M. Lilienfeld,

M.D., M.P.H., D.Sc. (Hon.), F.A.C.E.

New York

Oxford

Oxford University Press

1994

Oxford University Press

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland Madrid
and associated companies in
Berlin Ibadan

Copyright © 1976, 1980, 1994 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.,
198 Madison Avenue, New York, New York 10016-4314

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
Lilienfeld, David E.
Foundations of epidemiology.—3rd ed.
revised by David E. Lilienfeld, Paul D. Stolley.
p. cm.
Rev. ed. of: Foundations of epidemiology
Abraham M. Lilienfeld and David E. Lilienfeld. 2nd ed. 1980.
Includes bibliographical references and index.
ISBN 0-19-505035-5
ISBN 0-19-505036-3 (pbk.)
I. Epidemiology.
I. Stolley, Paul D.
II. Lilienfeld, David E. Foundations of epidemiology.
III. Title. [DNLM:
1. Epidemiologic Methods.
2. Epidemiology.
WA 950 L728f 1994]
RA651.L54 1994 614.4—dc20
DNLM/DLC for Library of Congress 93-5963

9 8 7 6 5 4 3

Printed in the United States of America
on acid-free paper

increase in magnitude if present before exposure; this response pattern should occur infrequently in persons not so exposed.

7. Experimental reproduction of the disease should occur more frequently in animals or humans appropriately exposed to the hypothesized cause than in those not so exposed; this exposure may be deliberate in volunteers, experimentally induced in the laboratory, or demonstrated in a controlled regulation of natural exposure.
8. Elimination or modification of the hypothesized cause or of the vector carrying it should decrease the incidence of the disease (e.g., control of polluted water, removal of tar from cigarettes).
9. Prevention or modification of the host's response on exposure to the hypothesized cause should decrease or eliminate the disease (e.g., immunization).
10. All of the relationships and findings should make biologic and epidemiologic sense.

Assessing Causality

The epidemiologist applies criteria of causality to the research before recommending clinical or public health actions. These criteria need not be satisfied in every way before causality can be inferred. Rather, they provide a framework for deriving a biological inference from epidemiologic and other scientific data. In practice, *a relationship is considered causal whenever evidence indicates that the factors form part of the complex of circumstances which increases the probability of the occurrence of disease and that a diminution of one or more of these factors decreases the frequency of the disease.* The etiologic factor need not be the only cause of the disease, and it may have effects on other diseases.

The following concepts are used by epidemiologists in making a causal inference:

- Strength of association
- Consistency of the observed association
- Specificity of the association
- Temporal sequence of events
- Dose-response relationship
- Biological plausibility of the observed association
- Experimental evidence

Strength of association

The strength of association is measured by the relative risk (or odds ratio estimate of the relative risk). A strong association between exposure and outcome

gives support to a causal hypothesis. When a weak association is found (for example, a relative risk of 1.2 to 1.5), other information is needed to support causality. Repeated findings of a weak association in well-conducted studies can still lead to effective public health action. When an exposure affects many people and the outcome is extremely adverse, a small increase in risk can be of major concern to public health officials. Action may be taken to lower the exposure and reduce the risk for large segments of the population. Strength of association supports a hypothesis of causality, but weak associations supported by other evidence of causality are sometimes equally important.

Consistency of the observed association

Confirmation by repeated findings of an association in case-control and cohort studies in different population groups and different settings strengthens the inference of a causal connection. Finding such consistency is logically equivalent to the replication of results in laboratory experiments under a variety of environmental or biological conditions.

Consistency of association can be illustrated by data from many studies of the relationship of oral contraceptives to cardiovascular disease. Many cohort and case-control studies have shown an increased risk of cardiovascular disease associated with oral contraceptive use in a variety of settings and population groups (Vessey, 1978).

Specificity of the association

It was formerly thought that to be causal, a one-to-one relationship should exist between the exposure and the disease; one exposure should cause one disease, and no other exposures should cause the disease. This has its roots in the bacteriological model where one microorganism is associated with one disease. In the study of chronic diseases, less emphasis has been given to specificity as a criterion of causality. The development of cancer is associated with a number of exposures, many of which are accepted as causal. Conversely, exposures such as smoking are associated with a number of adverse outcomes from cancer and cardiovascular disease to birth problems, and these associations are accepted as causal by the medical and public health communities. Specificity of a relationship between exposure and outcome strengthens confidence in a causal inference, but lack of specificity does not rule out causality.

Temporal sequence of events

It seems obvious that in order for an exposure to cause an event (disease), it must precede and not follow the disease. In many cases, the temporal sequence of events is clear-cut. One example is the study of prenatal exposures and malformations; it is usually easy to document that an exposure precedes the birth of

the malformed baby. However, for many other associations the temporal relationship is subject to debate.

In studying the relationship between age when breast-feeding ceases and infections of the baby, for instance, some researchers claim that longer duration of breast-feeding leads to fewer infections, but others claim that illness of the child leads to a cessation of breast-feeding. Which came first, the illness or the weaning? A cohort study design can resolve the issue of temporality, but for many study questions prospective studies are difficult or impossible to carry out.

Dose-response relationships

If a factor is of causal importance in the occurrence of a disease, then the risk of developing the disease should be related to the degree of exposure to the factor, i.e., a dose-response relationship should exist. The dose-response relationship between serum cholesterol level and the risk of coronary heart disease is an example. Another example is the relationship between duration of estrogen use and risk of endometrial cancer. Several studies also suggest that low-dose estrogen contraceptives carry a lower risk of venous thromboembolism than do higher-dose estrogens.

An observed dose-response relationship strengthens a causal hypothesis. Unfortunately, it is sometimes difficult to quantify an exposure in terms of a dosage or gradient. Dosage and duration of exposure are often interchanged in study designs, and both may cause a gradient in disease frequencies. Dosage can refer to the amount of a given exposure in a given time period, as in the number of cigarettes smoked per day, the amount of a hazardous chemical or particle in the work environment, or the amount of a drug taken each day. Information on actual dosage is often not available, so duration of exposure is substituted, as in years of cigarette smoking, years working in a given occupational environment, or length of time using a drug. Use of duration as a proxy for dosage necessitates an analysis that accounts for time; people with longer exposure times may have a greater time period in which to develop or discover the disease.

Biological plausibility of the observed association

A causal hypothesis must be viewed in the light of its biological plausibility. A causal association between ingrown toenails and leukemia, to take an absurd example, would be highly improbable. On the other hand, an association that does not appear biologically credible at one time may eventually prove to be so; indeed, the observation of a seemingly implausible association may actually represent the beginning of an extension of our knowledge. The established statistical association between circulatory diseases and oral contraceptive use is an excellent example of this. At first, there was no known physiological mechanism by which hormones could so profoundly affect the circulatory system. Yet, the statistical

association was present, and possible physiological mechanisms were later discovered, such as alteration of the clotting cascade, increased platelet adhesiveness, and direct effects on the arterial wall. It becomes important, therefore, to further investigate associations even if they are initially thought to be biologically implausible. The cigarette smoking-lung cancer relationship was initially considered biologically implausible by some, but carcinogens in cigarettes were identified, which lent biological plausibility to the observed association.

The ability to produce a particular disease in animals by exposing them to possible etiologic agents considerably enhances the causal hypothesis. Though one must be cautious in generalizing from the results of animal experiments to the human condition, this may be a relatively minor problem if the results of both animal experiments and epidemiologic studies in human populations are consistent. Animal experiments can also be valuable in determining the intermediate biological mechanisms that are involved in a disease, thereby providing the basis for seeking similar mechanisms in humans. Darwin's signal contribution to biological thinking was that the human species is not so unique a biological phenomenon as we may like to think; modern molecular biology confirms the unity of human and other animal species.

Experimental evidence

The randomized clinical trial (RCT) is the closest approximation in epidemiology to an experiment, and a well-run trial may confirm a causal relationship between an exposure and an outcome. The "exposure" is generally a drug, treatment, or procedure, and the outcome is reduction of disease or mortality. The Lipid Research Clinics Trial demonstrated that a pharmacological reduction in serum cholesterol led to lower heart disease, and other clinical trials have shown that pharmacologic lowering of blood pressure also reduces heart disease. Similar comments apply to the results of community trials. Ethics prevent the conduct of a trial of an exposure that is thought to have deleterious effects, and thus the randomized clinical trial and the community trial are limited to a subset of study questions related to potentially beneficial effects of an exposure.

Some situations approximate an experiment without the benefit of randomized, concurrent controls. The efficacy of inner-city comprehensive-care programs in reducing the incidence of rheumatic fever was demonstrated by comparing neighborhoods in a city that were similar to one another except for their eligibility for the programs, but the populations may have differed in ways not known or documented in the study (Gordis, 1973). Conversely, removal or reduction of an exposure may result in a decrease in disease. The decrease in smoking among physicians led to a decrease in lung cancer among physicians while rates in the general population continued to rise. The decline in the use of isoprenaline in England in the 1960s led to a decline in asthma-related deaths.

Exhibit C

16 Va. Envtl. L.J. 1

Virginia Environmental Law Journal

Fall 1996

Carl F. Cranor^a John G. Fischer^{aa} David A. Eastmond^{aaa}

Copyright (c) 1996 Virginia Environmental Law Journal Association; Carl F. Cranor, John G. Fischer, David A. Eastmond

JUDICIAL BOUNDARY DRAWING AND THE NEED FOR CONTEXT-SENSITIVE SCIENCE IN TOXIC TORTS AFTER DAUBERT v. MERRELL DOW PHARMACEUTICALS, INC.

I. Introduction	2
II. The Effect of Daubert v. Merrell Dow Pharmaceuticals, Inc.	6
A. Daubert v. Merrell Dow Pharmaceuticals, Inc.	6
B. Interpretations of Daubert	9
C. Problem Areas in Interpreting Daubert	14
1. The Distinction Between Methodology and Conclusion	14
2. The Distinction Between Admissibility and Sufficiency	16
3. The Standard of Judicial Review	18
III. The Threat of Overly-Simple Admissibility Rules	21
A. Epistemological Differences between Tort Law and Science	21
B. Simplified Admissibility Rules	25
1. Enshrining Daubert Considerations	27
2. Requiring Multiple Scientific Studies for Admissibility	28
3. Requiring Epidemiological Evidence	31
4. Special Restrictions for Interpreting Epidemiological Studies	32
a. Statistical Significance Rules	33
b. Relative Risk Rules	37
c. Sample Size and Duration of Studies	40
d. Using 'Hill's Factors' for Excluding Evidence	42
5. The Automatic Exclusion of Animal Evidence	49
C. Avoiding Temptations to Utilize Overly Stringent Admissibility Rules for Scientific Evidence	58
IV. An Alternative Account of the Admissibility of Scientific Evidence in the Law	62
A. Admission of Epidemiological Evidence	63
B. Admission of Animal Studies	65
C. Justification	72
V. Conclusion	76

*2 I. Introduction

What admissibility standards should govern the introduction of scientific evidence in toxic tort litigation? In many toxic tort cases, the only evidence available to prove that the toxic substance at issue caused the plaintiff's injury is scientific opinion testimony; thus, its introduction or exclusion may determine the outcome of the case. This question has remained open since the U.S. Supreme Court addressed this issue over three years ago, when it enunciated a new test for the admissibility of scientific evidence. The Supreme Court, in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹ rejected the Frye test² for the admissibility of scientific evidence in toxic tort cases, holding that it had been superseded by the Federal Rules of *3 Evidence.³ The Court wrote: "[I]n order to qualify as 'scientific knowledge,' an inference or

assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation--i.e., 'good grounds,' based on what is known."⁴ It further held that "Rule 702 . . . contemplates some degree of regulation of the subjects and theories about which an expert may testify."⁵ Thus, a trial judge "must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable."⁶ Moreover, there must be a "grounding in the methods and procedures of science," the knowledge "must be derived by the scientific method," and the knowledge must be "relevant" to the facts of the case.⁷ A ruling on admissibility thus entails a preliminary assessment "of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue."⁸

In light of the Daubert decision, this article addresses the issue of the use of scientific evidence in civil litigation, using the science required in toxic tort cases as an example. We aim not so much to provide an interpretation of the Daubert decision, but to address several substantive issues that the decision and subsequent interpretations raise about the role of scientific testimony in toxic tort litigation. The Court has now officially given judges a "gate-keeping" role⁹ in admitting scientific evidence which explicitly authorizes them to draw "boundaries" around what may be admitted as scientific evidence for tort law purposes. This is a power, however, that must be used sensitively and subtly to serve well the parties to tort litigation; this article seeks to provide some guidance as to how this power should be exercised.

When engaging in "boundary drawing," judges risk at least two generic mistakes: explanatory mistakes and strength mistakes. Judges make explanatory mistakes when they impose improper substantive restrictions on the content of evidence that may be admitted to prove that a substance causes human harm. For *4 instance, a judge might insist that a party provide either a particular kind of evidence, such as epidemiological studies, or multiple kinds of evidence, such as epidemiological, animal, and mechanistic evidence. Such restrictions would be mistakes because any appropriate scientific evidence that helps to explain the causal relationship between a defendant's actions and a plaintiff's injury is relevant and should be admissible. There are a variety of different explanatory paths that may lead to a conclusion; no particular explanation should be precluded as long as it has appropriate support. Moreover, many different legitimate scientists and scientific disciplines can contribute evidence in support of an explanation that a defendant more probably than not has harmed a plaintiff.¹⁰ Thus, as we discuss below, various proposed content restrictions on scientific explanations should be avoided.

A second generic problem is to mistake the strength of evidence required for a firm scientific conclusion that a substance causes human harm with the strength of evidence needed for a tort law conclusion. Tort law and scientific inquiry are different institutions, each with different evidentiary and social goals. A judge's failure to be sensitive to these differences when admitting scientific evidence in tort cases may inadvertently distort the law. We are concerned that judges may accept only certain kinds of scientific answers, given by certain kinds of scientists, to questions that are fundamentally legal in nature. In particular, judges might effectively change tort law standards of evidence, replacing them with scientific evidence standards more stringent than those that many respectable scientists would adopt. Judges should avoid this potential error, and should keep distinct the goals, mandates, and standards of tort law and science, in order to avoid mistaken admissibility decisions that may inadvertently change the desirable balance of interests between adversaries in tort litigation.

To address these concerns, we argue that courts should adopt admissibility rules that are sufficiently sensitive to allow the admissibility of all the evidence upon which scientists routinely rely to draw conclusions about harm from toxic substances. In this endeavor, courts should adhere to notions of admissibility in tort cases that reflect the goals and aims of tort law. Finally, courts *5 must maintain a fair balance of procedural and substantive interests between plaintiffs and defendants.

This article suggests considerations and standards for admissibility that attempt to balance the sometimes inconsistent goals of tort law and science. To introduce this discussion, in Section II we characterize the effect of the Daubert decision. Following a brief discussion of the facts of the case, we present and critique various views on the effect of the substance of Daubert. Those views advocate a range of interpretations, from overly liberalizing to stringently curtailing admissibility rules. In addition, we identify three procedural concerns raised by the decision which should similarly inform future admissibility decisions. In Section III, we then analyze some of the dangers posed by some types of admissibility rules offered in response to Daubert. In this section, we begin with the basic premise that evidentiary requirements for tort law and for scientific purposes are somewhat different. If these differences are not acknowledged by the courts they will inadvertently change the law. In addition to neglecting the epistemological contexts of tort law and scientific practice, courts may ignore the complexity of decisions made by scientists in their research. Such an approach risks the promulgation of “cookbook” admissibility rules--evidentiary rules that may appear as easy to use as the recipes in a cookbook--and the use of “cookbook” scientific evidence. That is, when courts are faced with the daunting task of evaluating the validity of scientific evidence, as Daubert commands, they may well develop, to change the metaphor, overly simple, “bright-line” criteria for the admissibility of scientific evidence that may be inappropriate in the tort law process. In this discussion, we identify and critique examples of such overly stringent admissibility rules. We then present a number of policy arguments against adopting such overly stringent rules in the tort law context. Finally, Section IV sketches an alternative view of how scientific evidence in toxic tort litigation might be addressed by the courts, focussing primarily on the use of animal studies, a particularly controversial area.

Generally, we feel the new focus on science after Daubert is salutary. However, because there is little guidance as to what constitutes admissible scientific evidence, there is a risk of erroneous application of the Daubert commands. On the one hand, courts risk excluding too much evidence. They must be more thoroughgoing in their acceptance of the wide range of scientific evidence--evidence routinely relied upon by the scientific community for arriving at scientific judgments. On the other hand, courts should ***6** be more sensitive to the mistakes that can arise from overly-simple views of scientific evidence--bright-line rules and overly-stringent evidentiary standards adopted from some understandings of science--and to the use of this evidentiary material in tort law. Ultimately, judges and lawyers may need to experience a quantum leap in understanding the subtleties of scientific inquiry in order to prevent various simplified views of scientific evidence from undermining and subverting the goals of tort law.

II. The Effect of Daubert v. Merrell Dow Pharmaceuticals, Inc.

A. Daubert v. Merrell Dow Pharmaceuticals, Inc.

The Daubert plaintiffs, Jason Daubert and Eric Schuller, were born with serious birth defects.¹¹ During pregnancy their mothers had taken Bendectin, an anti-nausea drug manufactured by defendant Merrell Dow Pharmaceuticals.¹² The minors and their parents sued Merrell Dow, claiming that Bendectin caused the boys' birth defects.¹³ Merrell Dow's expert submitted an affidavit that stated that no published study had found Bendectin to be a human teratogen and that therefore, use of Bendectin during the first trimester of pregnancy had not been shown to increase the risk of birth defects.¹⁴ Plaintiffs' experts concluded that Bendectin could cause birth defects, basing their conclusion upon: 1) test tube and animal studies linking Bendectin and malformations; 2) studies showing similarities between the molecular structure of Bendectin and other teratogens; and 3) a reanalysis of published epidemiological studies.¹⁵ The trial court, refusing to admit plaintiffs' evidence on the causation issue, granted summary judgment for Merrell Dow.¹⁶ The court held that, because there was a plethora of epidemiological evidence regarding Bendectin, plaintiffs' substantial non-epidemiological evidence was not sufficient to create a material issue of fact and defeat the summary judgment motion.¹⁷ The trial court relied on the Frye general acceptance test in its ruling.¹⁸

*7 The Ninth Circuit affirmed,¹⁹ also following the Frye general acceptance test.²⁰ The court apparently gave great weight to the fact that other appellate courts had not admitted reanalyses of epidemiological studies regarding the teratogenicity of Bendectin that had never been published nor peer-reviewed.²¹ Furthermore, it noted that the large number of published studies opposing plaintiffs' position that Bendectin could cause birth defects undermined the efficacy of reanalyses that reached the opposite conclusion.²² The appellate court concluded that reanalysis of epidemiological data is generally accepted in the scientific community only when there is sufficient peer review of such reanalysis.²³

The United States Supreme Court vacated the decision excluding the plaintiffs' scientific evidence and remanded the case for reconsideration under a newly enunciated standard for admissibility of scientific opinion evidence.²⁴ The Court held that the adoption of the Federal Rules of Evidence had superseded the Frye general acceptance test.²⁵ The Court noted that [Federal Rule of Evidence 702](#) spoke directly to the issue: "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise."²⁶

The Court considered the Frye general acceptance test "at odds with the 'liberal thrust' of the Federal Rules and their 'general approach of relaxing the traditional barriers to "opinion" testimony.'"²⁷ Thus, the Supreme Court certainly considered the new standard as more liberal with respect to the admissibility of scientific evidence than the prior Frye test, and that policy of more liberal admissibility represents the main thrust of the decision. The Court went beyond merely quoting the Federal Rule in announcing *8 a new standard for admissibility. Eschewing the Chief Justice's suggestion to restrict its opinion to the demise of the Frye test,²⁸ the Court offered, in dicta, some "general observations" concerning the proper standard of admissibility under the Federal Rules. The Court noted that [Rule 702](#) limited expert testimony to scientific knowledge.²⁹ In addition to having a certain aura of reliability, the proposed testimony must be relevant to the case at hand.³⁰ The trial judge must consider whether the methodology grounding the proffered opinion testimony is scientifically valid and whether it relates to the instant case in order for it to meet these two criteria.³¹ The majority opinion then outlined several non-exclusive factors for the trial judge to consider in evaluating whether the methodology is scientifically valid. These factors include: 1) falsifiability, or testability, of the theory guiding the technique used to reach the offered conclusion, 2) publication and peer review, 3) any known potential rate of error of the technique, and 4) general acceptance within the relevant scientific community.³² None of these factors should be treated as dispositive. The Court specifically noted, for example, that, for various reasons, some legitimate scientific methodologies may not undergo peer review.³³

Finally, in a somewhat puzzling passage, the Court addressed some concerns of the parties and amici on both sides of the case. The Court first dismissed the suggestion that abandonment of the general acceptance test would lead to a flood of "junk science" that would confuse juries. The Court noted: "Vigorous cross-examination, presentation of contrary evidence and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence."³⁴ In addition to this traditional protection, if the proffered evidence were truly of dubious value, the court could admit it but direct a verdict or grant summary judgment based on the insufficiency of the evidence.³⁵ The Court then considered the worry that gatekeeping judges, shackled by the chains of "scientific orthodoxy," would somehow stifle the search for truth.³⁶ Interestingly, the Court noted the differences *9 between the search for truth in the legal context and in the scientific context.³⁷ The Court pointed out that incorrect hypotheses are very useful in advancing scientific knowledge, particularly when their incorrectness is shown.³⁸ Such incorrect hypotheses, on the other hand, are of little use in the much quicker and more final context of a particular legal case.³⁹ The Court stated that a judge will occasionally incorrectly exclude valid scientific methodologies, but that such exclusion is part of the

balance to be struck in the legal context where the admission of an erroneous technique can have grave and irreparable consequences to the parties involved in an adversarial case.⁴⁰ All in all, the Daubert opinion is mixed, at times insisting upon the policy of liberal standards of admissibility, and at times emphasizing the “gatekeeping” role of a judge in excluding purportedly unreliable evidence.

B. Interpretations of Daubert

The Daubert opinion has generated a great deal of commentary, from courts attempting to apply Daubert in a particular case to the Defense and Plaintiffs' bars, whose respective members appear to interpret the opinion to support their particular point of view.⁴¹ The Defense bar has construed the opinion as a strong statement against the use of “junk science” in the courtroom.⁴² By contrast, the Plaintiffs' bar urges that the opinion requires liberal admissibility, so that more evidence will come in and be heard by a jury, where sympathy to victims may play a significant role.⁴³ Both views pose problems. A narrow interpretation of Daubert, in part engendered by the complicated task assigned to trial judges by that case's holding, risks the promulgation of overly-simplified admissibility rules, such as requiring the use of epidemiological studies to prove the toxic effects of a chemical. The use of such criteria risks excluding from juries toxicologically sound and relevant scientific *10 evidence that many scientists and scientific bodies find compelling in coming to their own factual conclusions.⁴⁴ Further, since plaintiffs have the burden of going forward with the evidence, the more demanding the criteria are for admitting scientific evidence, the greater the plaintiff's hurdles are in presenting its case to the jury. The most liberal view of admissibility, on the other hand, is probably precluded by the Daubert ruling itself. More mid-range views are represented by the collection of essays in the Federal Judicial Center's Reference Manual on Scientific Evidence,⁴⁵ which is a reference manual for federal judges on the state of scientific evidence in several fields pertinent to cases likely to fall under Daubert.

A more subtle and insightful approach than those proposed by advocates on either side and more theoretical guidance than that provided by the Reference Manual are both needed. Some perfectly good scientific evidence would be excluded under the most stringent admissibility interpretations, and some obviously invalid scientific evidence would be permitted under the most permissive interpretations. All parties to the debate need to recognize that there are different kinds and strengths of scientific evidence and that scientists differ in their assessment of the adequacy of the same evidence. Courts need to remember this in their interpretation and application of the principles in Daubert. Failing to understand these points may lead a court to decisions which endorse the kind of “scientific orthodoxy” that concerned the Daubert Court.

It is, of course, not surprising that members of the defense bar have attempted to portray Daubert as a strong judicial stand against “junk science.”⁴⁶ Marc S. Klein provides a forceful example *11 of this point of view.⁴⁷ He argues that Daubert resolved the previously unsettled issue of whether courts could function as gatekeepers with respect to scientific evidence.⁴⁸ Unfortunately, he cites only one case in support of the proposition that some courts have held that they could not properly fill the gatekeeping role.⁴⁹ Klein's conclusion suggests that Daubert represents a conservative view, meant to reign in the unfettered use of junk science in the courtroom.⁵⁰

Klein's strong conclusion exemplifies some of the more extreme interpretations of Daubert by the defense bar. Such a conclusion is clearly contrary to the Daubert opinion. The Court in Daubert, while acknowledging that judges should perform some sort of *12 gatekeeping function, emphasized the liberal thrust of the Federal Rules and stated that it was rejecting the more restrictive general acceptance test.⁵¹ Thus, Daubert was primarily a reaction against the overly restrictive Frye general acceptance test. Klein also states that “Daubert rejects the anti-intellectual, antiscientific argument that science is too demanding for our purposes and that, as a result, we should accept something less in the courtroom.”⁵² Klein here misconstrues the argument made by proponents of more liberal standards of admissibility.⁵³ In any event, the Supreme Court did not specifically draw the conclusion attributed to it by Klein. Rather, the Court

pointed out the difference between the search for scientific truth and the search for legal truth, thereby acknowledging that the legal standard should be different than the scientific standard.⁵⁴

Taking a different view, some Courts of Appeals appear to regard Daubert as liberalizing admissibility decisions. In *United States v. Posado*,⁵⁵ a criminal case, the Fifth Circuit Court of Appeals held that it could not establish, as it had in the past, a *per se* exclusion of a method or technology not generally accepted by the requisite scientific community.⁵⁶ Similarly, in *In re Paoli Railroad Yard PCB Litigation*,⁵⁷ the Third Circuit Court of Appeals emphasized the flexible nature of the inquiry by noting that a court should take into account any and all factors bearing on the question *13 of reliability rather than relying on bright line rules such as the general acceptance test or peer review.⁵⁸

Unquestionably, however, Daubert does not represent a complete liberalization of the standards for admissibility of scientific evidence. In dicta, the Court acknowledged that some valid scientific evidence may be excluded under the rubric of Daubert, thereby tacitly acknowledging that the traditional methods of challenging shaky evidence (for example, cross-examination, presentation of contrary evidence, or use of appropriate jury instructions) are not by themselves up to the task of preventing a jury from deciding a case based upon possibly faulty science.⁵⁹ The Supreme Court's designation of Rule 104(a), as opposed to Rule 104(b), as the proper standard under which to judge the validity of scientific evidence, also seems to bolster this view.⁶⁰ Rule 104(a) allows the court to determine the admissibility of evidence based on a preponderance of the evidence standard.⁶¹ Instead, the Supreme Court could have chosen Rule 104(b), which restricts a trial judge's examination of facts upon which the relevance of testimony depends to a determination of whether a reasonable trier of fact could find the fact to be true, a more liberal standard than Rule 104(a).⁶² This selection arguably suggests that the Supreme Court did not want to allow a jury to determine the admissibility of scientific evidence because a jury might not disregard evidence which it has heard but has determined to be legally irrelevant.⁶³ This choice of rule may also have been seen as guarding against the alleged proliferation of "junk science," used to confuse or influence juries.

Thus, in Daubert, the Supreme Court appears to have rendered an opinion liberalizing the standard for the admissibility of scientific evidence, but not adopting as liberal a standard as it could have. In addition to the dispute regarding the extent of the decision's *14 changes to rules of admissibility, the decision may also lead courts in another direction. Of great concern is the possibility that lower courts will interpret Daubert to support, contrary to the opinion, the use of bright-line rules, such as the four criteria noted in the opinion, to rule on the admissibility or inadmissibility of scientific evidence. The court did stress the flexibility of the trial court examination of the question, but, in assigning the daunting task of ruling on the admissibility of extremely complicated, technical, and sometimes novel scientific techniques, it may have tempted trial judges to look at a set of fixed criteria to handle this difficult procedure.

C. Problem Areas in Interpreting Daubert

In addition to the analysis of the substance of the Daubert decision, the opinion contains several procedural points which lower courts may misinterpret or find difficult to apply.

1. The Distinction Between Methodology and Conclusion

The Daubert Court carefully drew a sharp distinction between the scientific methodology and reasoning underlying scientific testimony and the conclusions drawn using the data at hand and the relevant methodology.⁶⁴ A trial court should only examine the reliability of the underlying methodology used in determining the admissibility of scientific evidence; it should not engage in an evaluation of the validity of the conclusions reached by an expert, the admission

of whose testimony is at issue.⁶⁵ The problem is that identifying the line between a methodology and the conclusions reached by employing that methodology is frequently difficult.

Moreover, consideration of different policies may influence exactly where a court decides to draw that line.⁶⁶ For example, in Wade-Greux v. Whitehall Laboratories, Inc.,⁶⁷ a district court excluded the plaintiffs' scientific evidence regarding the alleged teratogenic effect of an over-the-counter asthma medication. The Wade-Greux court held that, under the Daubert standard, to show that a substance is a human teratogen one must provide all of the following: 1) repeated, consistent epidemiological studies; 2) an animal study duplicating the defects; 3) a dose/response relationship *15 between the agent and the effect on the experimental fetus; and 4) a biologically sensible mechanism of teratogenicity of the agent.⁶⁸ Because the plaintiffs' evidence fell considerably short of these criteria, the court ruled much of the proffered scientific testimony inadmissible.⁶⁹ The court appeared to blur the methodology/conclusion distinction in its opinion by incorrectly examining the conclusions offered rather than the underlying methodology. It disagreed with plaintiff's experts that numerous cited epidemiological studies supported the conclusion that the agent does in fact cause birth defects.⁷⁰

When a court rejects scientific testimony on the grounds that the conclusions generated by the methodologies run counter to the conclusions of most other experts, it is rendering a decision based on the strength of the testimony. Consideration of the strength of the proffered evidence is proper in the context of deciding motions for summary judgments or judgments notwithstanding the verdict, where a court explicitly considers the weight of the conclusions. However, consideration of the strength of the evidence is improper in the context of ruling on admissibility. An interesting case for the methodology/conclusion distinction occurs where an expert cites a particular study as supporting her conclusions, when that study does not support those conclusions. This does not seem to implicate a question of admissibility, because the validity of the methodology is unquestioned. In *In re Paoli Railroad Yard PCB Litigation*, the court emphasized that a flaw in the expert's reasoning process does not involve a question of admissibility, unless the flaw is large enough to render the expert's reliance on an underlying study unreasonable.⁷¹ Thus, where the underlying methodology *16 meets the Daubert admissibility criteria, the court should not allow errors or gaps in the expert's reasoning from the underlying methodology or, more generally, conclusions drawn from the underlying methodology (however tenuous those conclusions may be) to render the expert's testimony inadmissible. The Paoli reasoning on this point is more sensitive than the Wade-Greux decision to both the range of scientific evidence and to the Daubert decision itself.⁷²

2. The Distinction Between Admissibility and Sufficiency

Daubert clearly addressed the issue of whether scientific evidence can be admitted, rather than the question of whether the evidence is sufficient to avoid a summary judgment or directed verdict and put the case before the trier of fact. To some degree, this distinction between admissibility and sufficiency is reflected in the methodology/conclusion distinction. The status of a particular methodology used to generate an expert's conclusions is evaluated by the court at the stage of possible admission of scientific evidence. If particular scientific testimony is deemed admissible, a court may still render a summary judgment or a directed verdict if the conclusions reached by the expert fail to establish a material issue of fact. In other words, when considering the sufficiency of a litigant's scientific evidence to prove a fact, the court may then look at the conclusions drawn by the expert based upon the scientific methodology. Of course, even at the admissibility stage, a court could consider the conclusions drawn by the expert to determine if such conclusions are somehow minimally grounded in the methodology, as well as if the conclusions are relevant to the fact the litigant seeks to prove.⁷³ In any event, it appears that courts *17 have excluded sound scientific evidence with some regularity in toxic tort cases, particularly cases involving Bendectin, as a means of disposing of such cases.⁷⁴ Further, courts sometimes explicitly employ sufficiency considerations, such as the presence of contrary evidence, to exclude evidence.

A difficulty here, of course, is that the courts may be disingenuous in disposing of cases on their merits in this fashion, thus encouraging standards for the admissibility of scientific evidence to develop in a skewed fashion. First, the courts may be underestimating the ability of juries to consider and resolve cases involving such complex scientific evidence. Second, they may exclude perfectly legitimate methodologies whose conclusions may perhaps fail to raise a substantial jury question, but whose admissibility seems clearly required under Daubert. The courts should handle that kind of sufficiency problem through a procedure such as summary judgment, directed verdict, or judgment notwithstanding the verdict if, in their considered view, it does not raise a material issue of fact to put before a jury. Confusion might arise in a situation where plaintiffs offer scientific testimony based on one epidemiological study suggesting that a particular chemical does cause the type of injury suffered by plaintiffs that is contradicted by numerous other studies. Perhaps because of the overwhelming number of contradictory studies, the presence of that one study may fail to raise a material issue of fact, such that defendants could win on summary judgment. However, that issue is one of sufficiency, not one of admissibility. The methodology may be sound even if the conclusion is dubious or one with which the judge disagrees.⁷⁵

*18 3. The Standard of Judicial Review

At this stage, it is unclear what standards of review federal circuit courts will apply when considering appeals of trial court rulings on the admission of scientific evidence. Most circuits have held that an abuse of discretion standard applies, such that a trial judge's ruling on the admissibility of scientific evidence must be "manifestly erroneous" or "clearly erroneous" before it can be overturned.⁷⁶ The Sixth Circuit has noted, however, that an appellate court will engage in de novo review of the issue of whether a trial court properly followed Daubert, since that is a question of law, not of fact.⁷⁷ The Third Circuit Court of Appeals, in *In re *19 Paoli Railroad Yard PCB Litigation*, addressed the judicial review of a trial court's exclusion of evidence. That court enumerated factors which a trial judge should consider when addressing the admissibility of scientific evidence, and it rejected a bare abuse of discretion standard of review of trial court rulings that exclude such evidence.⁷⁸

Paoli involved lawsuits by persons allegedly exposed to polychlorinated biphenyls (PCBs).⁷⁹ The trial court excluded some of the plaintiffs' evidence regarding causation.⁸⁰ The Third Circuit affirmed the trial court ruling on this evidence and then extensively discussed the application of the Daubert rubric. The court enunciated three factors in addition to the four set out in Daubert as relevant to the trial judge's screening task:⁸¹ 1) the "degree to which the expert testifying is qualified," 2) "the relationship of a technique to 'more established modes of scientific analysis,'" and 3) "the 'non-judicial uses to which the scientific techniques are put.'"⁸² Following Daubert, the court also emphasized that these factors are not exclusive.⁸³

Next, the court noted that the plaintiffs must make more than a *prima facie* showing of the reliability of the scientific methodology underlying proffered scientific testimony.⁸⁴ A proponent of scientific evidence must demonstrate by a preponderance of the evidence that the proffered expert opinion is reliable. However, the inquiry here is directed at underlying methodologies only, rather than conclusions generated by the use of such methodologies, however erroneous the conclusions might appear to the judge.⁸⁵ The *20 court then concluded that the distinction it had previously drawn between a methodology and its application is inconsistent with Daubert.⁸⁶

Finally, the Paoli court noted the tension between the Federal Rules' preference for admissibility and the traditionally extensive deference given to trial judges regarding the proper standard of review because of their superior vantage point on the evidence.⁸⁷ In response, the court stated that it would take a "hard look" at trial court decisions excluding scientific evidence as unreliable if that exclusion would result in summary judgment.⁸⁸ Thus, in toxic tort cases, where the exclusion of scientific evidence on causation almost always results in the inability of the plaintiff to prove the causation

element of his or her cause of action, the “hard look” standard would typically apply. This higher standard of review seems appropriate, for it takes into account the effect of the exclusion of scientific evidence on a litigant’s case and the procedural hurdles plaintiffs face in bringing a case before a court. The Paoli court implicitly acknowledged the notion that trial courts have excluded certain kinds of scientific evidence in order to prevent cases which judges consider non-meritorious from reaching the jury.⁸⁹

The foregoing procedural issues provide background and context for the main focus of our discussion of the admissibility rules for scientific evidence, the nature of the evidence, and the level of stringency that must be required for admission. As noted above, apparently courts have occasionally violated procedural distinctions that the Supreme Court argues should be preserved.⁹⁰ More frequently, courts have decided cases and commentators have urged decisions on grounds that exclude otherwise sound scientific evidence, impose too demanding evidentiary standards, or would distort tort law’s current balance of interests.

***21 III. The Threat of Overly-Simple Admissibility Rules**

A. Epistemological Differences between Tort Law and Science

A focus on the admissibility of scientific evidence may cause us to forget or overlook the fact that the epistemological contexts of scientific practice and legal practice are different. Such disparities, arising out of the divergent goals of scientific and legal practice, suggest the importance of using an approach that is sensitive to the institutional context for evaluating scientific evidence. In short, we need an approach for evaluating scientific evidence that is sensitive to the tort law context.⁹¹

To some extent the above epistemological concerns may echo aspects of the Daubert decision which caution against “allowing scientific assessments to intrude on the rights of parties to present evidence to the jury.”⁹² However, in order to be explicit about this point, we address this issue separately and not merely as an interpretation of Daubert. This larger issue is related to four substantive considerations that go beyond the particular focus on Daubert and interpretations of that decision.

First, to some extent, commentators and courts might be mistaken about conceptions of scientific evidence on which scientists themselves rely. Thus, much of what follows tests existing or proposed admissibility rules against examples and some of the evidentiary principles and practices utilized by scientists themselves. Realistic examples suggest a much wider range of scientific evidence than court opinions might or than many legal commentators sometimes use.

Second, courts may inadvertently adopt standards for admissibility that enshrine misleading standards of accuracy for tort law purposes. There are two kinds of incorrect outcomes for a tort law trial. In one case, a defendant may mistakenly be held accountable for injuries a plaintiff suffered (known in the technical language of science as a legal “false positive”). In the second case, a defendant *22 may be erroneously exonerated (known as a legal “false negative”), and a wrongfully injured plaintiff may go uncompensated. The tort law is equally concerned with avoiding both kinds of mistakes, and given the rules of tort law, this concern should be roughly equal between plaintiffs and defendants.⁹³

A problem, however, arises because of the interaction of scientific and legal standards of evidence. Scientific standards of evidence tend to be designed, or to have evolved asymmetrically, to *23 prevent false positive mistakes in science, with a lesser concern about false negatives.⁹⁴ If tort law unconsciously adopts the scientific concern with false positives as a result of its admissibility rules, it will adopt a mistaken conception of “accurate” or “correct” decisions for tort law purposes. Tort law has and should have procedural and substantive rules that have nearly equal and, ideally, relatively low numbers of legal false positives and legal false negatives.⁹⁵ Scientific inquiry, in contrast, aims to minimize false

positive mistakes. Nothing expressed in the Daubert decision appears to contravene this view. A sensitive approach to admissibility rules should avoid substituting the scientific concern for the overall tort law concern.

Third, we approach the admissibility of scientific evidence much as we would in a trial. Prior to the outcome of a trial, a court should be receptive to evidence or arguments regarding whether a particular substance causes a particular disease, such as cancer. One purpose of a trial is to determine whether, for tort law purposes, a substance is judged to cause a disease, whether the substance in question is judged to have caused the particular disease which plaintiffs have contracted, and whether defendants should be held accountable for the result.

The law provides institutional procedures to discover the legal truth about causation. A legal trial, however, is an instance of imperfect procedural justice.⁹⁶ We have a standard for assessing the correctness of the outcome of a trial independent of the procedures themselves, and the procedures do not guarantee a correct outcome. However, we should be careful about which standard we use to assess the correctness of the causation issues in a trial. Is the proper standard what scientists reasonably believe after taking into account all evidence available at the time of trial, or is it what they will ultimately come to believe about particular causal issues once all the information has been submitted? Some discussion *24 tends to suggest that the latter is the proper standard, and, unless evidence presented at the trial court tends to support such ultimate criteria, consideration by a court is problematic.⁹⁷ However, our view is that the appropriate standard should take into account all of the evidence available at the time of trial in determining whether causation is more probable or not. The outcome of a tort law trial with relatively liberal admissibility standards may fairly well approximate this standard.

Fourth, some proposed admissibility rules, either those used by courts or those recommended by commentators, may inadvertently undermine the procedural fairness of tort law. Evidentiary standards that are too demanding will impose a hidden factual burden of proof on plaintiffs that increases their procedural hurdles before they can bring their full case before a judge and jury. Such burdens for admissibility may be similar to or approach the criminal law's "beyond a reasonable doubt" burden of proof. This would distort tort law and upset the present balance of interests between plaintiffs and defendants.⁹⁸

Thus, our view is that different standards of evidentiary stringency are appropriate for different contexts of inquiry and for the varying goals of the particular institutions in question. While one standard of evidentiary stringency may be appropriate in the context of pure research science, another may be needed for the criminal law, and yet a third employed by tort law. Some courts and some commentators appear to give little recognition to this point.⁹⁹ *25 Given the liberal thrust of the Federal Rules of Evidence respecting admissibility, it seems that judicial and legal understanding of science must become more sensitive than some courts and commentators have allowed. Admissibility rules concerning scientific evidence must be tailored for tort law and its particular burdens of proof. We suggest the need for a more sensitive understanding of science, its limits, and its presuppositions in order to avoid the undesirable consequence of distorting the balance of interests between parties to tort litigation.

There are some trends in recent tort law decisions that, contrary to the letter and spirit of Daubert and contrary to good tort law policy, appear to preclude legitimate scientific evidence of causation in toxic tort cases. Some courts appear to require human epidemiological studies for a plaintiff's case to proceed beyond a preliminary hearing on admissibility of scientific evidence, and some appear to prohibit reliance on a combination of animal studies and short-term tests for evidence of carcinogenicity or teratogenicity.¹⁰⁰ Both approaches are contrary to good toxicology as it is currently practiced.¹⁰¹ Thus, we suggest that, as a result of Daubert, courts should embrace a wider range of evidence than some courts have considered to date, a range of evidence that is routinely utilized by the scientific community.

B. Simplified Admissibility Rules

The Daubert opinion stressed the need for a flexible set of criteria to determine the admissibility of scientific evidence. Nevertheless, it left the door open for, and perhaps even invited the use of, overly simple, “cookbook” admissibility rules. Because of the complexity of scientific issues, lower courts may shrink from the subtle but difficult task of evaluating and weighing the various kinds of scientific evidence for the context in question. Alternatively, courts may enshrine one or more of the criteria enunciated in Daubert as determinative, thus creating a bright-line standard with which to evaluate proffered testimony based upon a novel scientific methodology. For example, courts may focus on “general *26 acceptance” and peer review as determinative tests.¹⁰² They may not accept the validity of a relatively novel, but sound, scientific methodology where there may not be a strong acceptance of the methodology within the relevant scientific community. In addition, with respect to particular issues, such as the causation of a litigant's injury by an alleged teratogenic substance, courts may accept only certain methodologies to show scientific evidence of injury and exclude others, even if scientists would utilize a wider range of evidence. Courts may insist on multiple sources of evidence with a high degree of certainty for proof of causation when lesser evidence might adequately establish with a lower degree of certainty that there is sufficient evidence to survive the admissibility inquiry for purposes of tort law.¹⁰³ Or, relatedly, they might require for tort law admissibility that scientists be highly certain, as is appropriate according to the standards of their scientific discipline, that a substance causes cancer or birth defects instead of requiring that the evidence be sufficient for tort law admissibility purposes.¹⁰⁴

All of the above are mistakes, as illustrated in greater detail below. To remain faithful to the letter and spirit of Daubert, judges should avoid using cookbook rules. Some of the mistakes might be seen as matters of law, such as enshrining Daubert considerations as necessary conditions. Some of them might be seen as explanatory mistakes, such as an insistence that an explanation have a certain content or an insistence on human epidemiological studies. Additionally, some of them might be seen as mistakes as to the weight or the strength of the needed evidence.

Our view is that courts can and should admit sound scientific evidence of the kind scientists utilize to guide their judgments. Any appropriate scientific evidence that helps to explain the causal relationship between a defendant's actions and a plaintiff's harm should be relevant and admissible. Many different legitimate scientific disciplines and many different legitimate scientists can contribute evidence that a defendant has harmed a plaintiff. The court and the jury are entitled to hear from all of them. Moreover, judges should be sensitive to the strength or plausibility of evidence they demand for establishing tort law accountability. Courts *27 should acknowledge the considerably different evidentiary contexts of tort law compared with research science. Courts can inadvertently upset the normative balance of interests in tort law by requiring mistaken conceptions of adequate scientific evidence. Different legitimate scientific disciplines and different legitimate scientists may have substantially different evidentiary standards for judging that a substance causes a disease.

With regard to legal categories of evidence, our view is that plaintiffs must provide a scintilla of scientific evidence based on a reliable methodology to survive the admissibility review,¹⁰⁵ a somewhat greater amount of evidence (at least relative to evidence offered by the other side) to survive a sufficiency review, and, for the ultimate jury decision considering all the evidence, a sufficient amount of evidence to establish as more probable than not an explanation of the plaintiff's causation claims. However, to carry this explanatory burden does not mean, as we argue below, that only certain kinds of evidence are necessary or that a jury must be as certain as the most demanding research scientists would be in order to accept plaintiff's causation claim. None of these evidentiary showings, in our judgment, has to measure up to the very best evidentiary standards adopted in scientific fields. Nor do they need to have the same high degree of certainty that is required for a firm scientific conviction that a causal connection exists between exposure to a toxic substance and contraction of a disease.

1. Enshrining Daubert Considerations

Courts may focus on one element suggested by the Daubert Court, to the exclusion of others, thereby enshrining particular elements of the decision. In *Wade-Greux v. Whitehall Laboratories, Inc.*, despite its apparent deference to Daubert, the court appears to rely almost exclusively on the Frye general acceptance test.¹⁰⁶ Doing so flies in the face of Daubert, since the Supreme Court indicated in dicta that “general acceptance” by the profession was just one consideration among many (none of which was necessary) for establishing the scientific reliability of causal claims.¹⁰⁷ Plaintiffs’ *28 experts’ approaches might have satisfied other Daubert considerations, but the court did not address these considerations at all.

2. Requiring Multiple Scientific Studies for Admissibility

Wade-Greux illustrates another example of an overly-simple approach to admissibility. The plaintiffs argued that a mother’s use of Primate Tablets and Primate Mist, over-the-counter asthma medications sold by the defendant “caused TiaNicole Wade-Greux to be born with true malformation of her upper limbs and other skeletal defects.”¹⁰⁸ The trial court held that plaintiffs, in order to have their scientific evidence admitted, had to show that their claims about causation were supported by “repeated, consistent epidemiological studies; . . . an animal model that duplicates the defects resulting in the human from the exposure; . . . a dose/response relationship between the exposure and the effect on the experimental fetus; and . . . the mechanism of teratogenicity of the agent should be understood and make biologic sense.”¹⁰⁹ As we discuss below, most of the court’s necessary conditions are scientifically problematic. Requiring that all of these conditions be satisfied for admissibility seems problematic as well.

Although the *Wade-Greux* decision is not a leading one,¹¹⁰ it does illustrate the kinds of mistakes that need to be avoided. First, the court appears to have engaged in an assessment of the plaintiffs’ experts’ ultimate conclusions instead of their methodology, contrary to the law of the Daubert decision. Daubert suggests that the appropriate inquiry for admissibility concerns the soundness of each piece of an advocate’s evidence, not the total import of it.¹¹¹ Its total import should be addressed as part of a sufficiency review. Second, the court seems to take a literal textbook approach to admitting scientific evidence. Its approach seems to be that because standard toxicological or epidemiological references suggest that there must be epidemiological, animal, and other evidence *29 in support of the claim of causation¹¹² -- the combination perhaps considered singly necessary and jointly sufficient by some--before a firm substantive scientific judgment that a possible toxic agent is a teratogen can be made, courts should require all of this evidence before any part of it is admitted.¹¹³ Such multiple sources of evidence may be the best and ensure the greatest degree of scientific certainty. There are, however, several difficulties with this approach. First, the court appears not to assess the soundness and reliability of each piece of evidence, but to evaluate the entire package of evidence in support of plaintiff’s case. Second, the court places overly restrictive constraints on what is an appropriate explanation of plaintiff’s causal claims. The court appears to insist that certain categories of evidence be present, such as human, animal, and short-term mechanistic evidence, even though all of this evidence would not necessarily be required to explain with a preponderance of the evidence the causal relationship.

Even granting for purposes of argument the appropriateness of an approach that considers the adequacy of plaintiff’s total evidence, the court’s standards may be too demanding; they may require more than the amount of evidence which many in the scientific community would need to conclude that a substance is more probably than not a human teratogen. Such demanding standards should not be needed to survive an admissibility inquiry as described by the Daubert Court. A court adjudicating a tort claim need not be persuaded that it is a scientific certainty that a substance *30 is a teratogen; this required degree of certainty would substitute ultimate scientific burdens of proof for tort law burdens of proof. Thus, the *Wade-Greux* court appears to have violated the methodology/conclusion distinction articulated by the Daubert Court, evaluated the whole of the plaintiff’s evidence and not each piece of it, required a more constraining explanation

than necessary, and demanded the best or most certain evidence when adequate or good evidence is the most that should be needed to survive admissibility.

Daubert only requires that for evidence to be admissible, it must be “reliable” and “relevant.”¹¹⁴ Courts should assess what degree of certainty must be satisfied for each part of an adversary’s evidence to be scientifically “reliable,” yet it seems unlikely that it must possess the highest degree of certainty. In addition, much evidence will be “relevant” to explanations of causation even though it may not be determinative of the issue in question. Requiring that for plaintiffs’ evidence to be admissible it must possess all the features of the best and most certain evidence suggested by toxicological or epidemiological textbooks seems much too demanding. Third, as we will see, the Wade-Greaux court’s admissibility requirements on individual parts of the evidence appear directly to contradict the views of leading scientists on certain issues.¹¹⁵

Fourth, the court appears to require extensive toxicological and epidemiological evidence merely for admissibility. By contrast, Daubert notes that a scintilla of evidence is needed to survive an admissibility hearing.¹¹⁶ A scintilla of evidence would be well short of the best, ideal, or most certain evidence as judged by standard toxicology and epidemiology textbooks. The Wade-Greaux court appears to exceed greatly the scintilla minimum in imposing scientific requirements on plaintiffs’ experts. In short, the court, perhaps misunderstanding the evidentiary requirements of the field, places a much too demanding requirement on plaintiffs to survive the admissibility stage of the trial. Finally, the Wade-Greaux court appears to insist that statistical studies, including epidemiological studies, must be “statistically significant” at the .05 level as a necessary condition for admitting the study to show a causal relationship between a particular exposure and an increased risk of experiencing *31 a particular outcome.¹¹⁷ As we discuss below, not even experts in the field require such demanding statistical evidence for finding epidemiological studies helpful in understanding causation.¹¹⁸

3. Requiring Epidemiological Evidence

A third example of overly stringent admissibility criteria concerns both courts’ and commentators’ insistence on the necessity of epidemiological evidence for proof of a causal connection between a plaintiff contracting cancer and that plaintiff’s exposure to a possible carcinogen. A trend evident at least since the Agent Orange and Bendectin cases is a view by many courts that epidemiological studies are necessary for a plaintiff to prove causation in a toxic tort case.¹¹⁹

In the leading Agent Orange opinion,¹²⁰ Judge Weinstein stated that “[a] number of sound epidemiological studies have been conducted on the health effects of exposure to Agent Orange. These are the only useful studies having any bearing on causation.”¹²¹ Similarly, in *Lynch v. Merrell-National Laboratories, Division of Richardson-Merrell, Inc.*,¹²² the First Circuit Court of Appeals noted that non-epidemiological studies used “singly or in combination, do not have the capability of proving causation in human beings in the absence of any confirmatory epidemiological data.”¹²³ As one commentator has pointed out, this “implies that epidemiological evidence is a necessary prerequisite for a plaintiff to prevail.”¹²⁴ Other courts hearing Bendectin cases have come to similar conclusions.¹²⁵ Courts hearing other toxic tort cases have *32 concurred as well.¹²⁶ A few courts have resisted the impulse to enshrine epidemiological studies as necessary to prove causation.¹²⁷

However, as we briefly consider below, a sensitive understanding of the issues concerning the admissibility and relevance of epidemiological evidence clearly shows that scientific opinion testimony should not be required to rest on epidemiological evidence as a necessary condition for admissibility. Epidemiological evidence can be quite good evidence, and for robust, consistent studies even the best evidence, that a substance causes harm to humans.¹²⁸ But, it is not the only relevant evidence, nor even necessary evidence from a scientific point of view, for assessing the causal relations

between exposure to a substance and contraction of a disease such as cancer.¹²⁹ This issue is best seen when we discuss reliance on animal evidence, so we postpone consideration of it until later in this article.

4. Special Restrictions for Interpreting Epidemiological Studies

Apart from making epidemiological studies necessary conditions for plaintiffs' cases, some courts and commentators have gone further and have required or argued that the epidemiological studies must satisfy additional conditions before they can be admitted into evidence in toxic tort cases.¹³⁰ Some have insisted that such studies be "statistically significant."¹³¹ Others have insisted that the studies find a relative risk of at least two between the exposed and control populations.¹³² Still others have suggested that a list of factors *33 be satisfied by epidemiological studies before they can be admitted.¹³³

a. Statistical Significance Rules

Both before and after the Daubert decision, several courts and numerous commentators have insisted that epidemiological studies must be "statistically significant."¹³⁴ This requirement means that studies must have less than a certain low probability, typically below .05, that a statistical association between exposure to a toxic substance and a disease is not the result of random chance.¹³⁵ If studies do not satisfy this condition, they should be rejected as evidence in toxic tort cases. Typically, these courts would require, as the scientific community usually but not invariably does, that studies have five-percent odds (or less) of resulting in a positive association by random chance.¹³⁶ This requirement tends to be treated as something like a bright-line rule that epidemiological studies must satisfy for consideration by the legal system, even though many scientists, while recognizing its importance, do not necessarily regard it as determinative or decisive in judging issues of causation.¹³⁷ Thus, some courts and some legal commentators tend to regard statistical significance as a screening device for the admission of scientific evidence.

That such an approach is problematic can be seen from a variety of considerations. Some are reasons of science or interpreting scientific evidence and some are reasons of policy or philosophy. First, many within the scientific community itself are moving away from using rigid tests of statistical significance for interpreting epidemiological studies.¹³⁸ This was an issue between amici in the Daubert case,¹³⁹ but recent discussions suggest that the cutting edge *34 of the field seems to be moving away from tests of significance for two reasons: tests of significance are a kind of decision rule, useful for certain purposes but not others, and tests of significance reveal less about the data than other presentations of the evidence.¹⁴⁰

Thus, if tests of significance are treated as decision rules in legal cases, they should be designed for the specific context in question which argues against a uniform test of significance. It would also argue against necessarily using the same test of significance for research and many legal purposes, including tort law purposes. Moreover, it is in the interest of tort law to have evidence presented in the most informative manner possible, which again argues against strict and uniform tests of significance. Second, if scientific results are excluded merely because they are not statistically significant, one risks excluding important evidence and the decision might result in "far greater inaccuracy."¹⁴¹ The reason that this might result in greater inaccuracy is that demanding tests of significance asymmetrically prevent false positives, but permit more false negatives. And, "[p]reemptorily rejecting all studies that are not statistically significant would be a cursory and foolish judgment, particularly if there are multiple studies tending to show a consistent effect."¹⁴² Thus, court decisions might be more accurate on factual grounds if a wider range of epidemiological data were admitted. Because of this asymmetry the most demanding standards of scientific evidence may skew the outcomes of toxic tort cases in favor of defendants.¹⁴³ Further, there are policy reasons to be concerned about stringent statistical significance rules.

Statistical significance aims to keep the number of false positive results low in order to guard against the effects of random *35 chance.¹⁴⁴ A false positive occurs when one mistakenly identifies a substance as toxic, for example, as a carcinogen, when it is not.¹⁴⁵ By contrast, a false negative identifies a substance as not toxic, for example, as not a carcinogen, when it is.¹⁴⁶ The focus on preventing false positives by insisting on certain tests of significance has both scientific statistical implications and philosophical implications in tort law. Keeping the chances of false positives low, everything else being equal (including sample size and the relative risk one thinks it is important to detect), means that the chances of incurring a false negative will be higher.¹⁴⁷ In a statistical study when everything else is equal, one cannot reduce the chance of a false positive without increasing the chances of a false negative and vice versa.¹⁴⁸ The scientific point is that one cannot, without using very large sample sizes, have very low false positives, very low false negatives, and a study that will detect small relative risks, such as risks of two or smaller. Thus, there will be mistakes. Moreover, rigid rules of statistical significance requiring low false positive rates seem inconsistent with making the best assessment of causation based on the available evidence, because one kind of statistical error will always be favored.

A social and legal point emerges from the scientific one. Determining which mistake it is important to avoid for social and legal purposes is an important policy decision. In interpreting a statistical study, one can typically choose to keep the chances of a false positive or the chances of a false negative low, but not both (provided that the sample size is not large enough to do both and detect a relatively small relative risk).¹⁴⁹ Insofar as judges insist on tests of statistical significance that are low, such as .05 or less, which keeps the number of false positives low, as a matter of mathematics they will increase the odds of false negatives.

Keeping the chances of false positives low and making the chances of false negatives high, thereby making it more difficult to establish causation, greatly favors defendants and makes winning *36 more difficult for plaintiffs.¹⁵⁰ Thus, if the important point for tort law is that the chances of favoring the plaintiff should be about equal with the chances of favoring the defendant as a result of the rules governing the admissibility of scientific evidence, it is important to recognize that rigid rules requiring low statistical significance will systematically disadvantage plaintiffs. The greater the discrepancy between the chances of false positives and false negatives with the chances of false positives being lower, the more this disadvantages plaintiffs. Such a consequence appears seriously to distort the procedural rules of tort law.¹⁵¹ This consequence is even more worrisome if, as some suggest, juries and judges accept statistical evidence much less critically than other kinds of evidence. Statistical evidence is then given greater credibility and risks imposing particular hardships on plaintiffs.¹⁵²

We would do well to heed both a scientist and an academic lawyer on the issue of statistical evidence. Thirty years ago Sir Austin Bradford Hill posed the fundamental question scientists should ask themselves regarding causation: “[I]s there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?” He then proceeded to sum up the views of many scientists concerning statistical significance:

No formal tests of significance can answer those questions. Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects.

Beyond that they contribute nothing to the “proof” of our hypothesis.¹⁵³

Michael Green, an academic lawyer concerned about the inordinate influence of the Bendectin and Agent Orange cases, concludes a discussion of statistical significance as follows:

[T]he art of teasing out causal inferences in the absence of a mature epidemiologic record is far too complicated for courts seriously to review the methodologies and analyses involved. Making the ultimate

causal inference requires an *37 assessment not only of the quality of the epidemiology but the biological plausibility, based on what is understood about the mechanisms of toxicity. Indeed, one of the lessons of the Bendectin cases is that the courts are not truly engaging in greater scrutiny of experts' opinions; rather, they are adopting a few relatively simple screening devices. . . . Especially as the available universe of evidence gets thinner, inadmissibility decisions have significant risks.¹⁵⁴

In sum, rigid rules of statistical significance, while perhaps simplifying the job of screening epidemiological studies, risk ignoring salient scientific evidence, encouraging less accurate decision-making by failing to account for both legal false positives and legal false negatives, systematically disadvantaging plaintiffs, and thus upsetting the current procedural balance of interests between plaintiffs and defendants.

b. Relative Risk Rules

Still other courts have required that epidemiological studies must find a relative risk of at least two. The reason for this is that a relative risk of two indicates that twice as many people in a group exposed to a possible disease-causing substance contracted the disease as contracted the same disease in a control group. In such circumstances one can conclude that the diseased persons in the exposed group "more probably than not" had their disease caused by the substance, thus plausibly satisfying the ultimate tort law burden of persuasion on that issue. A number of pre-and post-Daubert courts and a number of commentators have endorsed the idea that an epidemiological study must reveal a relative risk of at least two in order for such evidence to be admissible.¹⁵⁵ This is quite problematic. Even more worrisome is the decision by Judge Kozinski, upon remand of Daubert to the Ninth Circuit Court of Appeals, that affirmed the trial court by focusing on the relevancy of the testimony of plaintiffs' experts.¹⁵⁶ Those experts could not testify that Bendectin more than doubled the likelihood of birth defects, because the highest relative risk to which one credible *38 expert could testify was 1.6-1.7.¹⁵⁷ They could only testify that Bendectin was capable of causing defects.¹⁵⁸ Therefore, their testimony was deemed not helpful to the trier of fact and inadmissible under Rule 702.¹⁵⁹ For courts to require that a relative risk be greater than two makes sense for purposes of the ultimate tort law burden of proof which must be met to persuade a jury, because, as Professor Green has put it:

In the absence of other information, any relative risk less than two would be inadequate to support a plaintiff's verdict. Thus, even a statistically significant study finding a relative risk of 1.7 should result in a directed verdict for defendant, in the absence of other evidence enabling a more refined assessment with regard to the plaintiff.¹⁶⁰

Even though Professor Green appears to have views similar to some presented in this paper, several remarks are in order. For one thing, rarely is other information absent. Thus, it is unlikely that an epidemiological study will have to stand as the only evidence on causation. Moreover, even if a study showing a relative risk of less than two is, by itself, inadequate to carry the plaintiff's ultimate persuasion burden, it surely should be admitted into evidence because it is relevant to the ultimate judgment of causation if there is any other information available. And, while plaintiff's case might or might not survive a sufficiency review, depending upon the other evidence presented by plaintiff compared with the kind and amount of evidence presented by the defendant, that is a judgment that must be made on a case-by-case basis when all such evidence is available, not decided *a priori* about one piece of evidence absent all others.¹⁶¹ Thus, it clearly appears to be an error to exclude epidemiological evidence simply because it reveals a relative risk less than two, unless there is no other supporting evidence. Even this may be problematic, as we discuss below.

*39 In support of the argument above, there is a striking example from radiation epidemiology that suggests it is a serious mistake to exclude epidemiological studies with relative risks less than two. Ionizing radiation has long been a known carcinogen. It appears to cause cancer in many human organ systems. One study shows that atomic bomb survivors from the Hiroshima and Nagasaki contracted leukemia and multiple myeloma, as well as cancer of the esophagus, stomach, colon, other segments of the digestive system, urinary tract, lung, lymph nodes, and a number of other sites.¹⁶²

Contrary to widespread belief,¹⁶³ the radiation exposures for many in these study populations were relatively low. What is striking about these findings is that epidemiological studies show, with 90% confidence, that all malignant neoplasms taken together except leukemia have a relative risk of less than two. Individual cancers which have a relative risk less than two include stomach, other parts of the digestive system, lung, and some other sites. Only leukemia, multiple myeloma, urinary tract, and colon cancer have relative risks greater than two. This finding is problematic for courts that have ruled inadmissible epidemiological studies with relative risks less than two. Since ionizing radiation is one of the best known carcinogens and one that scientists are certain causes cancer, courts, faced with plaintiffs who have been exposed to radiation, must either rule inadmissible epidemiological studies for all neoplasms for which there is not a relative risk of greater than or equal to two, or must admit the evidence and then engage in the sensitive and complex task of assessing and weighing the evidence that exposure to ionizing radiation caused the cancer at the site in question. Clearly, given the substantial evidence and degree of certainty about this carcinogen, the latter course seems much more defensible on both scientific and legal grounds.

Finally, requiring that epidemiological studies have a relative risk of two before they are admitted into evidence ignores a salient feature about such studies. A study that reports a relative risk of two from exposure to a toxic substance may disguise higher or low relative risks, on the one hand, resulting from higher or lower exposures respectively, or, on the other hand, from more or less *40 sensitive individuals respectively. For example, higher exposures might result in relative risks of four, while lower exposures might reveal relative risks of only 1.7. The weighted average of the different relative risks would yield an overall relative risk less than two. As scientists become increasingly aware of sensitive subpopulations, for example, resulting from genetic or other susceptibilities,¹⁶⁴ they may discover that average relative risks inadequately reveal the risk posed to sensitive subgroups. Tort law clearly protects sensitive subgroups,¹⁶⁵ but overly stringent admissibility rules might frustrate such protection. Thus, admissibility rules which preclude studies with overall relative risks less than two might prevent the admissibility of studies which included relative risks greater than two for individuals exposed to higher levels of a toxic substance or relative risks greater than two for particularly vulnerable groups. Clearly, the automatic exclusion of such evidence would unfairly disadvantage both biologically sensitive plaintiffs and those who were exposed to toxic substances at higher levels than in the study. Thus, epidemiological studies with relative risks less than two should not automatically be excluded from evidence.

c. Sample Size and Duration of Studies

Sample size and duration of epidemiological studies are topics which, while not specified by judges or commentators as restrictions on epidemiological studies (like statistical significance and relative risk, discussed above),¹⁶⁶ nonetheless merit a cautionary note. First, both can be shortcomings of a study. Epidemiological studies that are of too short a duration may fail to reveal an existing relative risk since the latency of the disease might be longer than the study. A similar procedural problem arises when epidemiological studies are based on too small a sample, since that may be inadequate to detect the true relative risk that exists. Second, the more stringently courts apply statistical significance rules and the smaller the relative risk between the disease rate in the exposed population and in the control population, the more the *41 shortcomings of sample size in epidemiological studies are exacerbated.¹⁶⁷

There is a complex relationship between sample size, the chance of a false positive, the chance of a false negative, and the relative risk a study is able to detect.¹⁶⁸ If a sample for the study of a relatively rare disease is small, even a well-designed and implemented study may have insufficient statistical power to detect a disease effect, even if one exists.¹⁶⁹ If researchers use a sample which is quite small in a study to detect relatively rare diseases, such as those typical of many cancers, and either the researcher or judge insists on less than a .05 chance of false positives, there is a risk of high false negatives or low statistical power. There is also a risk of being unable to detect relative risks of the disease that are relatively low, for example between 2 and 4.¹⁷⁰ Researchers are likely to report a “no-effect” result simply because their statistical tool for detecting it is too insensitive. This outcome is more likely depending on the degree to which researchers or judges insist on epidemiological studies with low chances of false positives.

Another problem judges and researchers should avoid is conducting studies which are of too short a duration. If subjects of a study are not followed for a sufficiently long period of time, a disease effect might be missed.¹⁷¹ Thus, a study of insufficient duration could also easily result in a judgement of “no effect” simply because the duration of the study was shorter than the latency period of a disease that might have resulted. This is particularly true of cancers which tend to have a latency period of five to fifty ***42** years.¹⁷² It is also likely that a longer study would be more sensitive to lower risks.¹⁷³

Judges need to be sensitive to both sample size and duration since either might lead to an erroneous “no effect” result. Such a “no effect” judgement would be the product of experimental design and would potentially present an inaccurate picture of the biological processes involved.

d. Using “Hill's Factors” for Excluding Evidence

Some courts and commentators have argued that epidemiological studies should not be automatically admissible, but should be required to satisfy further criteria. Many of these commentators have chosen to rely on a set of nine “aspects” of a statistical association between two variables, which were first proposed by Sir Austin Bradford Hill in 1965.¹⁷⁴ Hill said these factors should be considered in assessing whether the most likely interpretation of the relation between the variables is one of causal connection.¹⁷⁵ Several commentators have argued that even epidemiological studies showing a relative risk greater than two should not necessarily be admitted if they do not satisfy “Hill's criteria.”¹⁷⁶

***43** The first of Hill's considerations is the strength, or relative risk, of the association. However, he warned that by emphasizing the strength of an association, researchers should not “be too ready to dismiss a cause-and-effect hypothesis merely on the grounds that the observed association appears to be slight.”¹⁷⁷ Hill noted there are many examples in medicine where this is true, including the fact that, “[r]elatively few persons harboring the meningococcus fall sick of meningococcal meningitis . . . [and] [r]elatively few persons occupationally exposed to rat's urine contract Weil's disease.”¹⁷⁸

Hill also said scientists should examine the consistency of the relationship between the two variables. The inquiry is whether “it [has] been repeatedly observed by different persons, in different circumstances and times.”¹⁷⁹ Yet he noted that “there will be occasions when repetition is absent or impossible and yet we should not hesitate to draw conclusions.”¹⁸⁰

Specificity and temporality are the third and fourth of Hill's factors pertinent to evaluating the nature of the association.¹⁸¹ Specificity exists when an association is limited to a particular group of workers and particular sites or types of diseases.¹⁸² Hill particularly warned against over-emphasizing the importance of specificity, since, “[o]ne-to-

one relationships are not frequent. . . . In short, if specificity exists we may be able to draw conclusions without hesitation; if it is not apparent, we are not thereby necessarily left sitting irresolutely on the fence.”¹⁸³ Hill also recommended examining the temporal relationship of the association, asking “which is the cart and which the horse?”¹⁸⁴ Though this factor is included in Hill's list of non-exclusive factors, it has become recognized as the one factor which should actually be viewed as a requirement for the admissibility of studies.¹⁸⁵

***44** The fifth factor which Hill suggested is whether the association reveals a biological gradient or dose response curve.¹⁸⁶ For example, Hill stated:

the fact that the death rate from cancer of the lung rises linearly with the number of cigarettes smoked daily, adds a very great deal to the simpler evidence that cigarette smokers have a higher death rate than non-smokers. The comparison would be weakened, though not necessarily destroyed, if it depended upon, say, a much heavier death rate in light smokers and lower rate in heavier smokers.¹⁸⁷

In that case, Hill concluded, a more complex relationship would be needed to satisfy the causal theory.¹⁸⁸

Hill's sixth and seventh factors are plausibility and coherence.¹⁸⁹ While Hill said it would be helpful if the causation was biologically plausible, he was reluctant to require that this factor be met.¹⁹⁰ “What is biologically plausible depends upon the biological knowledge of the day. . . . In short, the association we observe may be one new to science or medicine and we must not dismiss it too light-heartedly as just too odd. . . .”¹⁹¹ Hill qualified the flexibility of the plausibility requirement somewhat by also including coherence as a factor. Hill said, “the cause-and-effect interpretation of our data should not seriously conflict with the generally known facts of the natural history and biology of the disease.”¹⁹² Hill used the example of arsenic, which “undoubtedly” caused skin cancer in humans, but had not yet been demonstrated to cause the same disease in animals.¹⁹³

Hill's final factors include experimental evidence and analogies. Hill said that appealing to experimental evidence is possible occasionally.¹⁹⁴ He also suggested judging an association by an analogy. For example, “[w]ith the effects of thalidomide and rubella before ***45** us we would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy. . . .”¹⁹⁵ Hill noted that these factors should be used to study associations before drawing conclusions about causation. “What I do not believe--and this has been suggested--is that we can usefully lay down some hard-and-fast rules of evidence that must be obeyed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*.”¹⁹⁶ What the factors can do is “help us to make up our minds on the fundamental question--is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?”¹⁹⁷

Some courts and commentators appear to interpret Hill's considerations as necessary conditions that must be satisfied for a statistical association to create a causal connection. One commentator, for instance, suggests that “[e]pidemiological evidence that meets none of [Hill's] criteria should be deemed automatically inadmissible under Rule 702.”¹⁹⁸ We agree this is the proper analysis, if it includes the temporality aspect, since it is clearly a necessary condition of cause and effect. However, if temporality is excluded, it is more problematic to say that if none of the remaining eight aspects is present, epidemiological evidence should be judged inadmissible.¹⁹⁹ Hill himself indicated that none of the other eight factors was necessary to be satisfied for an exposure to be judged a cause of a disease, since, for the remaining eight considerations, he cited examples in which (or reasons why) the consideration did not obtain, but causation did, thus

refuting the claim that all nine considerations must be satisfied before a conclusion of causation can be reached.²⁰⁰ In his preface to a reprint of Hill's essay, editor Sander Greenland notes:

*46 It is unfortunate that in the ensuing decades, this list or similar ones have been presented in textbooks as "criteria" for inferring causality of associations, often in such a manner as to imply that all the conditions are necessary. A careful reading of Hill shows that he did not intend to offer a list of necessary conditions; on the contrary, . . . he warned against laying down "hard and fast rules of evidence that must be obeyed before we accept cause and effect." As noted later Hill's only real mistake was to say that none of his nine aspects could be considered necessary if the association were indeed causal; in fact, temporality . . . is obviously necessary, as cause must precede effect.²⁰¹

In addition, the federal district court for the Southern District of New York, in *In re Joint Eastern & Southern District Asbestos Litigation*,²⁰² has evaluated evidence post-Daubert using Hill's considerations to argue that if the plaintiff's epidemiological studies do not satisfy any of Hill's considerations, then plaintiff's epidemiological evidence is not sufficient to survive a judgment as a matter of law following a jury verdict for the plaintiff.²⁰³ Again we would agree, if temporality is included, because it is the one appropriate necessary condition in the list. However, if the temporality factor is satisfied, the court's assertion is more problematic. First, the court converted Hill's considerations into criteria for judging the sufficiency of epidemiological evidence for legal purposes, an interpretation Greenland cautioned against and Hill himself disavowed. For example, Hill pointed out that there appears to be no question that in a nickel refining plant in South Wales the employees' exposure to nickel caused their lung and nasal cancer.²⁰⁴ He noted, however, there could be no repetition of the study because plant operating procedures had changed and animal studies had not confirmed the effects of nickel exposure.²⁰⁵ In short, Hill's consideration of consistency was not satisfied, but he appeared sure that *47 causation existed. Also, the increase in scrotal cancer in the nineteenth century among chimney sweeps exposed to soot was well known,²⁰⁶ but the consideration of plausibility was not satisfied because of insufficient medical knowledge at the time. Thus, contrary to the letter and spirit of Daubert, the *In re Joint Eastern* court is creating legal criteria even more stringent than considerations scientists themselves use.

The court's reasoning in *In re Joint Eastern* is also rejected by subsequent epidemiological articles, and by the recently published Federal Judicial Center's Reference Manual.²⁰⁷ Subsequent epidemiologists echo Hill's cautionary remarks about many of his factors. Rothman indicates that although strength retains "some meaning as a description of the public health importance of a factor . . . [it] is devoid of meaning in the biologic description of disease etiology" because whether an association is "weak" or "strong" depends upon the "prevalence of complementary component causes in the same sufficient cause . . . [T]his prevalence is often a matter of custom, circumstance or chance, and is not a scientifically generalizable characteristic."²⁰⁸ Other epidemiologists, *48 such as, for example, Mervyn Susser, echo Hill in expressing caution about insisting on relationships meeting the plausibility and coherence factors. "Coherence is an ultimate and yet not a necessary criterion for causality. . . . But coherence supports existing inference and theory."²⁰⁹ Susser continues: "[i]ncoherence may also have a more general explanation, in which instance it will generate a new theory. As Lilienfeld has said: 'the finding of a biologically implausible association may be the first lead to this extension of knowledge.'"²¹⁰

Regarding plausibility there is an implicit inconsistency if courts require Hill's factors to be satisfied, but rule animal studies inadmissible. Frequently, confirming animal and mechanistic studies are the best evidence that an epidemiological finding is biologically plausible. Some of the epidemiologists' sharpest criticisms are saved for those who would overemphasize the importance of "specificity." Some have noted that while there may be:

a tendency toward clustering of specific clinical features and other manifestations among patients afflicted with a particular cause of disease . . . and

. . . [it is possible to] find diseases in which there is very high association of a particular cause with a particular effect[,] . . .

....

. . . the majority of causal agents that are chosen as criteria for constructing disease entities are associated with a great diversity of clinical, pathological, and biochemical patterns. ²¹¹

Other commentators are more critical, claiming that:

[A]rguments that demand specificity are fallacious, if not absurd. There can be no logical reason why any identifiable factor, and especially an unrefined one, should not have multiple effects. . . . By now it is evident that the associations of health disorders with smoking depend on a variety of mechanisms, some causal and some not. Specificity *49 enhances the plausibility of causal inference, but lack of specificity does not negate it. ²¹²

Finally, the Reference Manual offers some guidelines on applying Hill's factors, noting that the temporal relationship must exist for causation. ²¹³ The Reference Manual also says that consistency with other research is an "extremely important factor," that biological plausibility provides "supporting evidence," but epidemiological evidence that is not implausible "should not be disregarded" because some disease processes are better understood than others, and that alternative explanations and confounding factors "should be examined and ruled out to avoid reaching an erroneous conclusion." ²¹⁴ The Reference Manual then notes that it is "never possible to rule out every alternative explanation." ²¹⁵ Lastly, the Manual indicates that, for specificity of association and dose-response, while strengthening the inference of causation, "absence of either does not weaken the inference," which is explicitly contrary to Bernstein's argument. ²¹⁶

In sum, some courts use (and some commentators recommend) rules for admissibility that are more stringent than leading scientists themselves would use. Moreover, insistence on all or most of Hill's criteria would erect evidentiary barriers that scientists themselves would not use. This could possibly preclude the use of sound, relevant evidence and impose additional procedural hurdles on litigating parties. Hill's considerations bear on the strength or the weakness of the evidence; except for the temporality consideration, they are not decisive criteria for rejecting it.

5. The Automatic Exclusion of Animal Evidence

Many courts exclude animal evidence, unless it is accompanied by epidemiological evidence, as relevant to judgments about the causal connection between exposure to a substance and contraction of a disease. An important decision by Judge Weinstein, *In Re Agent Orange Product Liability Litigation*, ²¹⁷ has influenced a number of courts to exclude animal studies per se from evidence in *50 toxic tort suits. ²¹⁸ Apart from what appear to be errors in an understanding of toxicology, this exclusion seems mistaken on two counts. First, it appears contrary to the Daubert Court's emphasis on the consideration of scientific evidence that is relevant to causation. Second, if one believes that a context-sensitive science is appropriate for courts to consider in assessing the toxicity of substances and their effect on human beings, then such evidence should not be excluded automatically.

In *In re Agent Orange*, Judge Weinstein argued that, “the studies on animal exposure to Agent Orange, even Plaintiffs’ expert concedes are not persuasive in this lawsuit. . . . There is no evidence that plaintiffs were exposed to the far higher concentrations involved in [the animal studies]”²¹⁹ Weinstein said because the animal studies involved different biological species, they were not helpful to the case.²²⁰ He said the studies “are of so little probative force and are so potentially misleading as to be inadmissible. . . . They cannot be an acceptable predicate for an opinion under Rule 703.”²²¹

While Judge Weinstein placed an emphasis on “this lawsuit,” his opinion has been widely interpreted as excluding reliance on animal studies, unless they are accompanied by epidemiological evidence.²²² Excluding reliance on animal studies, even in the absence of human epidemiological studies, seems mistaken, however. Although readily apparent differences exist between laboratory animals and humans, such as size, lifespan, and metabolic rate, from a biological or biochemical point of view there are also a large number of important similarities.²²³ For example, the biochemical and metabolic processes carried out in most organs are similar,²²⁴ although the observed rates of metabolism may differ.²²⁵ As a consequence, in the majority of cases close relationships can be seen in the responses of humans and laboratory animals to toxic *51 and carcinogenic agents.²²⁶ David Rall, while serving as head of the National Institute for Environmental Health Sciences, noted, “there are more physiologic, biochemical, and metabolic similarities between laboratory animals and humans than there are differences. These similarities increase the probability that results observed in a laboratory setting will predict similar results for humans. Clearly the accumulated evidence in the field of carcinogenesis supports this concept.”²²⁷ Moreover, of the approximately forty individual chemicals that have been recognized by the International Agency for Research on Cancer (IARC)²²⁸ as cancer-causing agents in humans, every one for which there is adequate data in experimental animals has been shown to be carcinogenic in animals.²²⁹ There is generally a close correspondence between the target organ in humans and at least one of the animal species studied.²³⁰ Researchers have also observed a similar response between humans and animals for other types of responses to toxic agents.²³¹ For the interpretation of animal data and its extrapolation to humans, however, adjustments and scaling factors need to be applied to account for differences in body size, surface area, lifespan, metabolic rate, or other differences that may exist between species.²³² Occasionally effects have been seen in humans where a similar response has not been seen in animal studies.²³³ Such cases, *52 however, tend to be exceptions rather than the rule.²³⁴ Thus, for toxicologists, the fact that there is information from “other” biological species is both relevant and helpful evidence. Moreover, contrary to Judge Weinstein, such studies have considerable probative force even if they might not always be as strong on evidentiary grounds as thorough epidemiological studies.

In fact, animal studies have considerable probative value for toxicologists, and this should be reflected in the law. Other scientific and legal bodies utilize animal studies in reaching conclusions about toxicity. The similarity in response of living organisms to toxic substances forms much of the basis for predictive and regulatory toxicology, and is relied upon by the Food and Drug Administration (FDA), the Environmental Protection Agency (EPA), the Consumer Product Safety Commission (CPSC), the Occupational Safety and Health Administration (OSHA), and other federal and state agencies for establishing allowable exposure levels for the safe usage of drugs, cosmetics, pesticides, disinfectants and other household and industrial products.²³⁵ Consequently, testing of chemicals using in vitro systems and experimental animals is generally required by these various agencies.²³⁶ While some agency assessments of the risks from toxic substances are forward-looking and preventive in nature, their deliberations are quite relevant to tort law considerations. That is, tort law often finds itself in the same position as scientific investigators trying to construct a causal explanation of what caused a death or serious disease. For this, toxicologists utilize all toxicological information bearing on the causal claim in question.²³⁷ For both preventive purposes and retrospective accountability in tort law, toxicologists act as detectives to try to identify causal paths that might be harmful (in a preventive case) or were harmful (in a tort case).

*53 Not all agency deliberations are predictive and preventive in nature. The FDA and parts of the EPA are required by law to evaluate substances before they enter the market (acting under so-called “pre-market approval statutes”) and before there is any significant exposure.²³⁸ Toxicological evaluation of substances in these circumstances is more predictive and explicitly preventive in nature. However, other agencies are not so predictive and preventive. A number of regulatory bodies, for example OSHA and other parts of the EPA, act under post-market statutes, and thus must act as scientific investigators and reconstruct a causal explanation of what lead to disease or death, taking protective steps as a result.²³⁹ These scientific inquiries are much more like those of tort law; therefore, the conclusions and deliberations are even more pertinent to tort law inquiries than are those of agencies engaged in strictly predictive and preventive toxicology.

Thus, while Judge Weinstein may be correct (and toxicologists agree with him) that good human data is the best evidence for making toxicological inferences about the effect of substances on human beings, it is not the only helpful data. Other data should not be preemptively judged inadmissible because it is from another animal species or not thought to have “probative value.” Non-human evidence clearly can and does have considerable explanatory or probative value for toxicologists in absence of epidemiological studies and it should be considered in toxic tort cases. Non-human data also importantly supplements or casts doubt on human data. For example, animal data can rule out a positive epidemiological study as having no biological plausibility. By contrast there are a number of substances identified as possible or probable human carcinogens on the basis of animal or mechanistic studies by the National Toxicology Program (NTP) and IARC, even though there are no or inadequate epidemiological studies.²⁴⁰

Fortunately, however, Weinstein's *In re Agent Orange* opinion may be quite limited. There are special considerations present in that litigation which may make his views an exception to a more general rule which favors admitting all evidence relevant to judging issues of causation in human beings. At the time of the *In re Agent Orange* litigation, there were governmental epidemiological studies showing no serious adverse long-term health effects from exposure *54 to Agent Orange.²⁴¹ Further, conflicting epidemiological studies were considered to be either inapposite or flawed.²⁴² Even so, such contrary evidence really does not bear on the admissibility issue, but could support a summary judgment.²⁴³

Recently, other courts have recognized the limitations of Weinstein's views.²⁴⁴ They have ruled that, as a matter of scientific practice, animal studies are the kind of evidence on which scientists rely for evidence of causation from toxic substances. Of particular note is a recent decision from the Third Circuit Court of Appeals, which discussed the mixed state of case law on admissibility of animal studies and noted that “[m]any cases have held that the studies are admissible.”²⁴⁵ The court added:

*55 While other cases have held that animal studies are inadmissible, these cases are for the most part distinguishable because most involved the exclusion of animal studies in the face of extensive epidemiological data that failed to support causation, because none involved studies on animals particularly similar to humans in the way they react to the chemical in question, and because none involved studies the federal government had relied on as a basis for concluding the chemical was a probable health hazard [as was true in this case].²⁴⁶

The Paoli opinion seems based on much better ground than some of the opinions noted earlier, since toxicologists often rely upon animal studies for evidence of causation from toxic substances.²⁴⁷ However, even the Third Circuit's view of the pertinent evidence may not be wide enough. That is, given what toxicologists know and how they view the evidence

they regard as pertinent to making causal judgments, courts should be open to a wider range of toxicological evidence than even the Third Circuit suggests. It is not clear that animal studies should be excluded even in the face of extensive epidemiological evidence to the contrary. Such evidence might or might not pass a sufficiency review,²⁴⁸ but that is a separate matter. Moreover, animal evidence may well be pertinent to judgments of toxic substances causing human harm (depending upon the content of that evidence) in the face of mixed epidemiological studies or no epidemiological studies.

An interesting counter-example to a claim about the irrelevance of animal evidence is provided by a scientific detective story. A Centers for Disease Control scientist described a case study in *56 which animal evidence, combined with other circumstances, led to a discovery of two deaths and to the criminal conviction of the person who was responsible for poisoning them with dimethylnitrosamine.²⁴⁹ The suspect, the spurned lover of the victim's wife, had spiked lemonade in the victim's refrigerator with dimethylnitrosamine, a yellow, water-soluble substance that causes severe liver damage.²⁵⁰ The spiked lemonade caused five people to become sick and two to die from liver necrosis. The suspect, an employee at a cancer research institute, desired to cause cancer in the victim's family to watch them die slowly, but had chosen a compound that was also acutely toxic.²⁵¹ The compound was also quickly metabolized, making it difficult to trace.²⁵² Investigators from the Centers for Disease Control were able to rule out other liver-damaging agents because they are not as toxic.²⁵³ They found one feature of dimethylnitrosamine which permitted it to be detected in forensic analysis: it causes methylation of the nucleic acid bases of DNA, such as guanine, which can be detected and measured by high pressure liquid chromatography tests.²⁵⁴

The important point about this example is that there was little or no prior human data showing these toxic effects. Virtually all the toxicological evidence came from animal or in vitro studies.²⁵⁵ Moreover, this was a criminal case with a higher burden of proof than tort law. Thus, if some of the rules concerning the non-admissibility of animal evidence in tort cases had been applied to exclude the evidence in that criminal case, a criminal would have gone free. More important, the scientists used all of the toxicological evidence they had available to them, most of it based on mechanistic and animal studies, to solve the crime. Data about mechanism, carcinogenic doses and lethal doses came from animal or in vitro studies, not human epidemiological studies. Of even greater interest is that this case was solved using essentially a case report--typically disavowed *57 by courts in toxic tort cases²⁵⁶--and animal studies along with other non-human toxicological data.

This example raises more problems for the view that human epidemiological evidence is necessary to establish actual causation in toxic tort suits. Some reviewers of this article have indicated that the circumstances establishing the defendant's causal connection to the deaths were so unusual,²⁵⁷ thus ruling out any other explanation, that toxicological evidence was unnecessary.²⁵⁸ While we tend to disagree with that analysis,²⁵⁹ it does suggest a deeper point about establishing causation in such cases.²⁶⁰ What is needed to establish causation in a tort case is an explanation that is more probably than not true,²⁶¹ connecting the defendant's actions to the plaintiff's injuries. However, providing an appropriate explanation does not automatically require the use of only human epidemiological evidence. Neither does it mean that epidemiological evidence is a necessary condition of such an explanation or that all forms of toxicological evidence relevant to establishing claims are necessary conditions for each explanation.²⁶² Only an appropriate explanation is necessary to establish causation. Thus, judicial and commentator insistence that the explanations have certain necessary components is mistaken. While human epidemiological evidence can be very good evidence, it is not necessary; the preceding example shows how a plausible, in fact powerful, explanation for injuries *58 to plaintiffs can be established without it. Moreover, much like the case just discussed, there are cases in which animal evidence conjoined with short-term test and structure-activity relationships might well be sufficient to show more probably than not that a substance is a human carcinogen.²⁶³

In conclusion, some courts and commentators have suggested enshrining into law more stringent criteria for judging the validity of scientific inferences and explanations than are required in the science itself. The courts are not being faithful to the science and the use of rigid rules may lead to consideration of a narrower range of evidence than scientists themselves would evaluate. The examples show there may be a variety of causal explanations for the same conclusion; which one is plausible in a particular case depends upon what evidence is available and what it shows. Legal failures to be sensitive to the subtlety of scientific inference also risk skewing and distorting legal relationships between plaintiffs, who have the burdens of proof to establish factual claims and to remove uncertainty, and defendants, who benefit from uncertainty. Defendants benefit from uncertainty because, if sufficient uncertainty is not removed from the plaintiff's evidence, then the plaintiff loses. The more a defendant can show that there is uncertainty or unanswered questions about plaintiff's evidence, the better for the defendant. Courts need to monitor such issues sensitively and carefully in order to avoid inadvertent prejudice to the interests of the parties.

C. Avoiding Temptations to Utilize Overly Stringent Admissibility Rules for Scientific Evidence

In light of the above discussion, it seems important to avoid the temptation to adopt overly stringent admissibility rules for scientific evidence for several reasons. First, using such rules would be contrary to the Daubert decision, which emphasizes the importance of scientific evidence. Second, using overly demanding rules would result in a cavalier rejection of perfectly sound, albeit sometimes not the best or most pristine, evidence. The result would be to exclude evidence which is clearly relevant to decisions about causation, that is, it would place mistaken substantive constraints on causal explanations. Third, to adopt stringent admissibility or sufficiency rules would be inconsistent with scientific practices which *59 emphasize using a range of evidence from many sources to make decisions about causation.²⁶⁴ Fourth, respectable scientists in different scientific fields, and even scientists within the same field, often disagree about how much evidence is adequate to support a judgment about a causal property of a particular substance. For example, a number of prominent epidemiologists believe the evidence between exposure to electromagnetic fields and the development of cancer is sufficient to warrant further studies into the effects of this agent.²⁶⁵ In contrast, other respected scientists, most notably biophysicists, claim the association between electromagnetic fields and cancer has "no persuasive scientific basis"²⁶⁶ and that resources should be directed towards more meritorious or pressing areas.²⁶⁷ Similar disagreements between well-respected scientists can be found for the health effects of other agents like asbestos,²⁶⁸ lead,²⁶⁹ and 1,3-butadiene.²⁷⁰ Moreover, within toxicology, differences arise between practitioners regarding the amount of evidence needed to justify claims of causation.²⁷¹ In sum, epidemiologists *60 clearly have differing opinions about the adequacy of evidence for a judgment of causality.

Consider the following discussion between two epidemiologists on the effect of confounding factors in epidemiological studies. Sometimes researchers have evidence that a substance, for example cigarette smoke or asbestos, harms human health, but continue to search for possible confounders to explain away observed associations between exposure to the substance and contraction of disease.²⁷² This delays action and frustrates health protection. One scientist, Sander Greenland, argues, "One can always invoke unmeasured confounders to explain away observational associations. Thus, actions should not depend on the absence of such explanations, for otherwise action would never be taken."²⁷³ Another scientist, an advocate of a careful search for confounders in such circumstances, H. J. Eysenck, tries to establish such causal connections with "proof in the sense usually accepted in science," or possibly proof "beyond a reasonable doubt," because such facts if discovered will slay "a beautiful hypothesis."²⁷⁴ Greenland and Eysenck clearly have distinctly different motivations for their beliefs. Greenland has a greater concern to protect public health, so is less willing to delay decisions indefinitely out of a desire to reduce uncertainty.²⁷⁵ Eysenck, on the other hand, may be seeking obviously sufficient proof to justify a highly certain scientific inference of a causal connection.

The discussion in the preceding two paragraphs illustrates the point that two researchers, with access to the same evidence, can reach quite different conclusions about causation. Courts should not apply such stringent admissibility or sufficiency rules that they do not allow, as a matter of law, a legitimate scientific point of view. Such admissibility and sufficiency rules threaten to legally foreclose otherwise open scientific issues and disputes. These issues should be left to triers of fact and not decided by hard and fast admissibility rules. Judges drawing boundaries around scientific evidence should not be seduced by either a univocal standard of evidence (which might enshrine evidentiary standards of one *61 discipline to the exclusion of another or include only the most demanding standards of evidence for a particular discipline) or stringent standards of evidence designed for the world of scientific research. Rather, judges should use standards which serve the legal purposes of tort law and recognize the legitimacy of a wide range of evidentiary standards held by respectable practitioners of all disciplines, in order to provide for better informed judgements about toxic effects in toxic tort cases.

A fifth reason to avoid overly stringent admissibility rules is that they risk enshrining a misconception of both science and what constitutes good and relevant evidence on contested factual issues. These rules risk “freezing in” or “freezing out” particular views of a particular field of science, like toxicology or epidemiology, while it is developing. It also risks freezing in place particular philosophies of science, a field currently in considerable flux. ²⁷⁶

A sixth reason is that some of the admissibility standards are likely to enshrine high admissibility barriers and thereby threaten to preclude evidence until it is established beyond a reasonable doubt. For example, one of the authors of this Article argued in another publication that the adoption of certain statistical rules of significance may risk such results because of the mathematical interaction between tests of statistical significance, sensitivity of tests, relative risks for which one can test and sample sizes. ²⁷⁷ Requiring low statistical significance when a study has a small sample size mathematically can force high rates of false negatives and result in skewed evidentiary requirements. ²⁷⁸

Finally, placing too great a burden on the admissibility or sufficiency of scientific evidence hides important policy issues behind the science. These issues include the following: who should bear the risk of harm resulting from disease more likely than not caused by exposure to toxic substances; when should wealth be shifted; and whom should decide these decisions, a judge or jury. These policy issues should be addressed on their own merits, not decided through proxies such as debates about scientific evidence. Juries' consideration of reliable scientific evidence is a legitimate debate. *62 However, that debate often overshadows other important substantive positions. There is a wide range of legitimate scientific disagreement, especially on the “frontiers” of scientific knowledge of the kind likely to surface in toxic tort cases, with reputable scientists on both sides of the debate. The law should make room for these other contested issues and not preclude one side or the other from testifying because of admissibility rules.

IV. An Alternative Account of the Admissibility of Scientific Evidence in the Law

In the preceding Parts we argued that there are a number of shortcomings in the responses of courts and commentators to the recent Daubert decision. In closing we sketch an alternative view of how courts might address scientific evidence in toxic tort litigation, focusing on one of the most contested areas: the use of animal studies. Other authors have addressed issues about the relative benefits of the use of epidemiology versus the use of other kinds of scientific evidence. ²⁷⁹ Short-term tests, *in vivo* and *in vitro* alike, and structure-activity tests are somewhat more distant from the evidentiary needs of tort law. All of these kinds of evidence are scientifically appropriate and relevant to judgments of causality and none should be ruled out of court by admissibility rules. We do not discuss them further in this Article, however.

We submit that all evidence on which scientists rely when making judgments of causality should be admissible in toxic tort cases, including: epidemiological studies, animal studies, case-studies, structure-activity relationships, and other short-term tests. This assertion is based in part on Daubert's emphasis on the use of scientific evidence based upon valid

methodologies,²⁸⁰ and in part on the need to find only an appropriate explanation to causally link defendant's actions and plaintiff's injuries.

The rules for admitting scientific evidence in tort law should preserve the traditional balance of interests between parties to a dispute and the traditional goals of tort law: to compensate victims for the harmful conduct of others which more likely than not harmed the victims and to deter others from engaging in conduct that will probably harm others. Admissibility rules should not explicitly or implicitly change the burdens of proof so dramatically that plaintiffs must establish a piece of scientific evidence to a very *63 high level of certainty, approaching the criminal law's "beyond a reasonable doubt" burden of persuasion, to satisfy admissibility conditions. Courts should not demand that each piece of scientific evidence on which expert testimony is based satisfy the very best or most certain scientific evidentiary standards in order to be admissible.²⁸¹ Courts must recognize the following: the difference between appropriate but minimally needed evidence for establishing tort causation and the "most certain" evidence; that scientists can and do hold different opinions about the kind and amount of evidence needed to make a causal inference that a substance more likely than not causes harm to humans; and that scientists differ on what is minimally adequate evidence in support of claims. Plaintiffs should be required to provide the following: a scintilla of reliable scientific evidence to survive the admissibility review²⁸²; a somewhat greater amount of evidence, at least relative to evidence offered by the other side, to survive a sufficiency review;²⁸³ and a sufficient amount of evidence to establish more probably than not plaintiff's claim of causation to actually win the lawsuit. None of these evidentiary showings, in our judgment, necessarily needs to measure up to the very best evidentiary standards adopted in scientific fields, nor do they need to have the same high degree of confidence that is required for a certain scientific conviction that a causal connection exists between exposure to a toxic substance and contraction of a disease. Finally, if existing rules of tort law are unfair to plaintiffs or defendants, these should be explicitly addressed and changed, rather than modified by subterfuge through overly strict, and possibly ad hoc, admissibility rules, like those in the Bendectin cases.

A. Admission of Epidemiological Evidence

Other authors have addressed the need for a more sensitive use of epidemiology in the law, so we merely highlight some of those *64 considerations without arguing for them in detail. Generally, epidemiological evidence can be the best evidence on the causal effects of a substance on human beings. However, its admissibility should not depend on whether it is the best and most certain scientific evidence. Scientists weigh all the available evidence in order to make inferences about causation and courts should not do less. Courts should even consider epidemiological evidence that is clearly not the best one could have, if it is the kind of evidence scientists would weigh in the balance.

Thus, courts should not always insist on strict and low statistical significance rules, a relative risk of at least two (at least when there is other supporting evidence), or a requirement that all or most of Hill's factors²⁸⁴ be satisfied before admitting scientific evidence. These issues bear on the strength or weakness of the epidemiological evidence to be admitted; this is a matter of degree. Failing to be statistically significant at a low level increases the odds that the study will be a false positive as a result of random chance. Failing to find a relative risk of two decreases the chance that plaintiffs will prevail before the triers of fact, unless there is supporting evidence. Likewise, failing to satisfy Hill's factors, assuming the temporality factor is satisfied, weakens the case for causation, but none of these by themselves decisively defeats plaintiffs' claims. However, a study strengthens plaintiffs' cases if it satisfies conventional statistical significance rules, finds a relative risk of at least two, and satisfies all of Hill's factors. It is important to notice that these considerations are just that--"considerations." They bear on the degree of strength of the case, but are not by themselves, or collectively, decisive reasons for rejecting or admitting evidence. We have tried to illustrate this fact by reference to the scientists themselves who regard these evidentiary considerations as matters of degree. Thus, in light of Daubert, it seems courts must be as sensitive as scientists are to the subtlety, complexity, strengths, and weaknesses of scientific evidence, and not issue overly simple rules for admitting or barring available evidence.²⁸⁵

Our point should not be misunderstood, however. Our argument is not that judges should become scientists. Rather, they should not erect artificial evidentiary barriers which will preclude legitimate, respectable scientists who may disagree with one another from testifying about the evidence for causation in a toxic *65 tort case. We try to provide some sense of the complexity of the issues involved so that judges and lawyers can begin to recognize with some degree of subtlety the myriad sources of evidence that are available and that can be relied upon in such cases.

B. Admission of Animal Studies

Obvious biological or biochemical differences exist between laboratory animals and humans. However, there are also a large number of important similarities. As a consequence, in the majority of cases, as we argued above, close relationships can be seen in the responses of humans and laboratory animals to toxic and carcinogenic agents. Typically, there is also a close correspondence between the target organ in humans and at least one of the animal species studied.²⁸⁶ A similarity between humans and animals for other types of responses to toxic agents has also generally been seen.²⁸⁷ Of course, there are and will be exceptions, but the failure of toxic responses to be universal between humans and animals should not undermine the judgment that in the majority of cases there is a similarity of response. What is needed for tort law is that the probabilities favor similarity of response, not that it is always similar between animals and humans or conversely. Occasionally effects have been seen in humans where a similar response has not been seen in animal studies. Such cases, however, tend to be exceptions rather than the rule.²⁸⁸ The biological and biochemical responses are so similar between animals and humans and sufficiently reliable that environmental and occupational regulatory agencies utilize them for establishing allowable exposure levels for the safe usage of drugs, cosmetics, pesticides, disinfectants and other household and industrial products.²⁸⁹ As noted earlier, one consequence is that testing of chemicals using in vitro systems and experimental animals is required by numerous regulatory agencies.²⁹⁰

*66 In many cases, the adverse effects of chemical agents were identified in animals before similar effects were seen in humans²⁹¹ or would likely have been seen had animal testing been required. In some cases, animal test information was used to protect large numbers of people from birth defects, cancer and other toxic effects. In others, the experimental and epidemiological data were either ignored or intentionally suppressed, resulting in serious medical conditions and, in some cases, the deaths of exposed individuals.²⁹² Examples include thalidomide,²⁹³ 1,2-dibromo-3-chloropropane (DBCP),²⁹⁴ and asbestos.²⁹⁵

*67 There are circumstances in which certain toxic effects have been seen in animals in which no counterpart has been seen in humans.²⁹⁶ However, since adverse effects seen in animals are not always seen in humans, regulatory scientists and researchers have developed procedures to evaluate animal data and its relevance to humans.²⁹⁷ Generally, this involves a weight-of-evidence approach in which all of the relevant data is evaluated. This is the procedure adopted by IARC²⁹⁸ and by the NTP.²⁹⁹ Human data, animal test results, data from in vitro test systems, similarities in chemical structure with other known toxic or nontoxic agents, and mechanistic information are evaluated. In addition, the quality of the study, the strength of the association, the results from other studies in related species, and the potential for bias and other confounding factors are also considered. Based on the overall weight of the evidence in assessing potential carcinogens, IARC classifies an agent as definitely carcinogenic to humans (Group 1), probably carcinogenic to humans (Group 2A), possibly carcinogenic to humans (Group 2B), not classifiable as to its human carcinogenicity (Group *68 3), or probably not carcinogenic to humans (Group 4). Similar classifications are used by the NTP and various regulatory bodies.³⁰⁰

Although human epidemiological data is generally required for a chemical to be classified as a Group 1 carcinogen, supportive data from related studies can play an important role. For example, ethylene oxide (ETO) was recently

classified as a Group 1 carcinogen.³⁰¹ In this case the epidemiological evidence was not conclusive with some studies showing elevated cancers of the lymphatic and hematopoietic systems whereas others failed to show any increase. However, based on the strength of the supporting data, ETO was classified as a definite human carcinogen.³⁰² In this case the supporting evidence indicated that ETO is a direct acting alkylating agent that has been shown to induce genetic damage and bind to proteins and DNA in the blood cells of animals and exposed workers. This agent has also been associated with lymphatic and hematopoietic tumors in animals and has been shown to induce mutations and chromosomal damage across all species.

The significance of the ETO example for tort law should be clear. Even though there is mixed epidemiological evidence, the weight of evidence from animal studies, various short-term tests, and mechanistic information was judged sufficient in the face of that evidence for IARC to classify ETO as a known human carcinogen. Two observations seem apt. First, some tort courts in ETO toxic tort cases might well have found the epidemiological evidence inconclusive, and thus excluded plaintiff's studies, and then excluded all the other evidence because there was no supporting epidemiological evidence. Yet, this would clearly be a mistake on scientific grounds and under the Daubert ruling. Second, since classifying a substance as a known human carcinogen requires *69 meeting more stringent evidentiary standards than tort law's ultimate burden of persuasion, a plaintiff ought to be able to bring a suit for ETO-caused injuries on the basis of the same evidence.

IARC also categorizes chemicals as to their carcinogenicity in experimental animals. In the preamble to its monographs, IARC notes that "all known human carcinogens, studied adequately in experimental animals have produced positive results in one or more animal species."³⁰³ In addition, they state that "[a]lthough this association cannot establish that all agents and mixtures that cause cancer in experimental animals also cause cancer in humans, nevertheless, in the absence of adequate data on humans, it is biologically plausible and prudent to regard agents for which there is sufficient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans."³⁰⁴ A similar statement could be made for most other types of toxic effects. An important issue is the exposure concentration or dosage at which these effects would occur in exposed humans. By reducing exposure to levels well below those at which no adverse effects are seen in animals and through an understanding of the mechanisms underlying the toxic effects, the probability of adverse effects occurring in humans can be minimized.

The critical issue in using animal evidence seems to be the following. One inquiry that courts need to conduct concerning carcinogens, for example, is whether, based on the evidence available, a substance can be judged more probably than not to cause cancer in human beings. A second inquiry is whether, on the basis of the evidence available, the substance in question can be judged more probably than not to have caused plaintiff's particular injuries. With respect to the second issue, clearly the courts need to be presented with evidence that connects plaintiff's exposure to a toxic substance with plaintiff's injuries. Even very good animal studies (or epidemiological studies for that matter) will not be on point to establish this second issue. However, if animal studies are sufficient proof from a toxicological perspective, they should be able to support a claim that a certain substance more probably than not can cause cancer or some other disease in humans. This is confirmed by approaches taken by toxicologists and scientific bodies such as IARC and NTP. IARC makes the judgment that something *70 is a "known" human carcinogen, or that it is a "probable" human carcinogen.³⁰⁵ It seems much too demanding to require that for tort law purposes that a substance must be judged to be a "known" human carcinogen in the IARC sense before scientific evidence can be admitted. Moreover, IARC's criteria for judging that a substance is a "probable" human carcinogen seems quite appropriate for tort law and its burdens of proof. Given the stringent NTP and IARC criteria for judging that something is a probable human carcinogen, tort law judges may justifiably take the additional step of saying that NTP or IARC establishment of a substance as a probable carcinogen makes a *prima facie* case for causation in humans.³⁰⁶ The Third Circuit Court of Appeals in Paoli recognized that "the 'more probable than not' standard employed by EPA [for classifying carcinogens] is the same standard that is employed in civil litigation."³⁰⁷ In addition, given the NTP and IARC criteria for evaluating whether there is sufficient

evidence to judge that a substance is a carcinogen in animals, it is plausible to use this for admissibility in tort law litigation. After all, in the absence of adequate human data, IARC regards animal studies as providing biologically plausible evidence that a substance is a human carcinogen.³⁰⁸ Thus, if, on the basis of animal evidence “it is biologically plausible and prudent”³⁰⁹ to regard that substances for which there is sufficient evidence for carcinogenicity in animals “as if they presented a carcinogenic risk to humans,”³¹⁰ tort law should not dismiss such evidence. Such an assessment of particular substances would then have to be evaluated to see whether it is sufficient to carry the ultimate tort law burden of persuasion. The outcome of such an assessment would depend upon how “plausible” the “risk” is and what the “risk” is thought to be.³¹¹ However, *71 it is surely biologically relevant to the judgment of causation, and should not be automatically dismissed.

An example of a substance that makes this point is 4,4'-methylenebis (2-chloroaniline) (MBOCA). There are “no adequate epidemiological studies of workers” exposed to MBOCA.³¹² One epidemiological study in the United States is in progress, but as noted earlier, as yet has insufficient duration (because of the long latency of bladder cancer)³¹³ to be sensitive to lower risks. A second study in Great Britain found excess risks of cancer from exposure to MBOCA, but “work records were insufficient to identify workers employed directly on the MBOCA production process.”³¹⁴ Other evidence, however, including animal studies, structure-activity relationships, and mechanistic information, indicates that MBOCA is likely to be a human carcinogen.³¹⁵ Moreover, our conjecture is that a large majority of toxicologists would likely regard MBOCA as a human carcinogen. What does a court do, faced with a tort law claim for worker injuries suffered as a result of exposure to MBOCA? It seems to us that this is the case in which animal and mechanistic evidence as well as structure-activity associations surely should be admitted on behalf of a plaintiff. Additionally, based on this evidence it seems more probable than not that MBOCA can cause human bladder cancer. IARC has listed MBOCA as a “probable human carcinogen,”³¹⁶ and the NTP has listed it as “reasonably anticipated to be a human carcinogen.”³¹⁷ Whether a particular exposure might cause a particular plaintiff’s bladder cancer would depend upon the facts of the case. Thus, it seems such evidence for MBOCA’s carcinogenicity should be judged admissible; it also goes a long way toward establishing a case for such a hypothetical plaintiff.

In considering the admissibility of animal evidence we concur in Green’s judgment that:

[W]hen epidemiologic evidence is lacking, thin, of questionable validity and ultimately inconclusive, dismissing other toxicological evidence is unjustifiable. . . . [P]laintiffs should be required to prove causation by a preponderance of the available evidence, not by some predetermined standard that *72 may require nonexistent studies. . . . [A court should consider] the universe of available evidence of toxicity.³¹⁸

C. Justification

The justification of the above views rests on four substantive considerations. First, courts should not adopt admissibility rules that preclude scientific evidence on which scientists themselves routinely rely; we have tried to indicate the problems with existing or proposed admissibility rules by means of examples throughout this Article. Second, courts should adopt evidentiary standards that give due consideration to the notion of tort law accuracy in decisions; that is, tort law should provide roughly equal protection to avoiding both legal false positives and legal false negatives. Courts should not demand “accuracy” in the scientific sense which aims to control only the false positives. Third, the test of accuracy in causal judgments for the court should not be whether, sometime in the future when all the facts are in, a court made the correct decision; rather, it should be whether, taking into account all the evidence available at the time of trial, plaintiff’s case more probably than not is favored by the scientific evidence. This is a generalization of Green’s point above. Fourth and finally, courts’ admissibility rules should not inadvertently undermine the balance of interests in tort

law--something we have argued a misconception of scientific evidence could easily do--rather they should preserve the balance of procedural and substantive interests between the parties.

First, throughout we have argued by means of examples that courts should not adopt admissibility rules which preclude consideration of sound scientific evidence on which scientists themselves rely. This is clearly a mistake which future courts should rectify. Some, such as the Third Circuit Court of Appeals decision in *In re Paoli* appear to recognize this issue.³¹⁹ The *Paoli* case appears to be a decision in the right direction.

Second, courts should adopt evidentiary standards that give due consideration to the appropriate notion of tort law accuracy in decisions; that is, tort law should provide roughly equal protection against both legal false positives and legal false negatives. This standard is what is suggested by the ultimate burden of persuasion in tort law--plaintiffs should have a slightly greater chance of *73 losing (because they have the preponderance of evidence burden of persuasion) than defendants, but their chances should be nearly equal. Thus, "accurate" decisions in tort law should take into account the possibility of both legal false positives and legal false negatives. This is contrasted with the primary concern in science to avoid false positive mistakes. If courts were to demand "accuracy" in the scientific sense which aims to minimize false positives, this would substitute the scientific notion of accuracy for the tort law notion.

Third, we approach the admissibility of scientific evidence much as we would in a trial. That is, prior to the outcome of a trial, a court is undecided as to whether a particular substance causes a particular disease, such as cancer. One purpose of a trial is to determine whether, for tort law purposes, a substance causes a disease and whether the substance in question caused the disease in question to the particular plaintiffs. Thus, we view the legal truth about causation *ex ante* before there has been an appropriate legal inquiry. Moreover, the law provides institutional procedures to discover the legal truth about causation. A legal trial, however, is an instance of imperfect procedural justice. That is, we have a standard for assessing the correctness of the outcome of a trial independent of the procedures themselves, and the procedures do not ensure a correct outcome. A trial results in a decision that may or may not be correct as judged against an independent standard. That independent standard for causation would be the body of scientific evidence and what it indicates about claims of causation. However, one should distinguish between what scientific evidence ultimately might show about causation and what scientific evidence might show at the time of trial about the causal issues. That is, at the time of trial a well-informed, wise and impartial scientist reviewing the evidence might or might not agree with the judgment about causation resulting from the trial.

More worrisome, however, is that the scientific evidence at the time of trial might be so mixed or uncertain that an impartial scientist could not come to conclusions about causation, just as a trial court might find it difficult to come to such a conclusion. A court, however, must decide the issue one way or another; it does not have the luxury of postponing judgment as a scientist might. However, we should be careful about which standard we use to assess the correctness of the causation issues in a trial--should the standard be what scientists know or reasonably believe at the time of the trial or instead, what they will ultimately come to believe about *74 particular causal issues when all the information is in and evaluated to the satisfaction of the most demanding scientific standards of evidence? Some discussion tends to suggest that the latter is the proper standard and unless evidence presented at the trial court tends to support such ultimate criteria, some might claim that it is problematic for a court to consider it.

However, our view is that the appropriate standard should be what it is reasonable for scientists to believe at the same time as the trial taking all evidence at the time into account. This proposal cuts both ways, sometimes favoring plaintiffs, sometimes favoring defendants. On the one hand, some substances once were thought not to be human carcinogens, e.g., asbestos and 1,3 butadiene, but subsequent studies have shown them to be known (asbestos) or probable human carcinogens (1,3, butadiene). 1,3 butadiene, for example, based on mechanistic information and evidence from animal and human studies is now believed to be a human carcinogen. Had a plaintiff brought a case that 1,3 butadiene caused cancer too early, plaintiff should have lost based on the information known at the time. Now the balance of the evidence might tend to favor plaintiffs. On the other hand, a substance such as 1,4-dichlorobenzene in the past was considered

to pose risks of kidney cancer to humans based upon increased incidence of cancer in male rats. More recent evidence suggests that this was probably a mistake because 1,4-dichlorobenzene interacts with a rat kidney protein not present in humans; thus, it plausibly is not a human carcinogen. ³²⁰

The point of these examples is that sometimes courts in tort cases will make mistakes on the basis of the evidence available, sometimes favoring one side, sometimes favoring the other. The test for whether they have made a mistake should be what the balance of available evidence at the time of the trial indicates. However, we do want to caution against using a different standard. Some scientists have argued that before one can claim that a substance is a carcinogen on scientific grounds one must have multiple epidemiological studies, multiple animal studies, numerous short-term tests, and all the results must cohere with one another. We might think of this as the ultimate and most demanding scientific ^{*75} standard for assessing evidence. While in a perfect world with perfect information this might be correct, there are several problems with this. For one thing, such extensive information is available only for a very few substances. Thus, this is a nearly impossible standard to meet for all but about twenty-five or thirty substances. Moreover, because of the paucity of information about most substances, this would severely and systematically disadvantage plaintiffs, even if they might have good, but not the very best evidence in support of their case. It would also asymmetrically advantage defendants who could merely point to the absence of proof on one of the tests and win their case. Finally, there might be good but incomplete evidence that would show more probably than not that a substance was a human carcinogen and had caused plaintiff's cancer. This evidence would suffice for plaintiff's case, but would be precluded by such a stringent standard. Faced *ex ante* with scientific evidence about a new substance and with claims that it caused harm to plaintiffs, a court only has its own procedures and processes to determine, based on the available evidence, causation issues. Because there are likely to be competing views of causation, there will be controversies. Each side will present its "picture" of the causal issues with the evidence that is available. The concern of this paper has been to argue against bright line rules that can inadvertently deprive one side or the other of legitimate evidence which might otherwise be in its evidentiary palette for presenting a picture of the causal connections. The justification for having a context-sensitive approach to the admission of evidence is to ensure a kind of procedural fairness, and perhaps, secondarily, to ensure more correct outcomes. The two are related in important ways.

On the one hand, ensuring that all the relevant evidence is admitted increases the chances that there will be a correct outcome on the factual issue. We argued above that some rules of admissibility adopted by courts and some suggested by commentators would preclude toxicologically sound scientific evidence. If such evidence is excluded, this raises the risk of mistaken decisions. Recall the preceding point that tort law must be sensitive to two kinds of mistakes: legal false positives and legal false negatives. More sensitive admissibility rules will help ensure this result.

On the other hand, adopting some of the court rules or those recommended by commentators will make evidentiary standards too demanding and thereby inadvertently undermine the procedural fairness of tort law. Demanding that litigants meet evidentiary ^{*76} standards which are created for research purposes will impose a hidden factual burden of proof on plaintiffs that increases their procedural hurdles before they can bring their full case before a judge and jury. Such burdens for admissibility may be similar to or approach the criminal law's "beyond a reasonable doubt" burden of persuasion. This would appear to distort tort law and upset the present balance of interests between plaintiffs and defendants. Surely such an unintended consequence should be avoided by tort law.

Fourth and finally, adopting scientific notions of evidentiary stringency or scientists' notions of accuracy will have inadvertent normative outcomes--they will inadvertently undermine the balance of interests in tort law. We have indicated throughout how this can occur. Roughly, as some of us and others have previously argued, demanding scientific standards of evidence, as represented, for example, by stringent rules about statistical significance, greatly favor defendants because they protect against false positives, but, unless quite large samples are the object of study, will increase the chances of false negatives (which will disadvantage plaintiffs). If courts inadvertently or deliberately adopt research science standards of evidence or inadvertently or deliberately adopt a concern to prevent scientific false positives, both

of which greatly advantage defendants, this will clearly upset the balance of interests between plaintiffs and defendants in tort law and undermine the fairness of the present system.

V. Conclusion

The Daubert decision focused attention on and requires valid and reliable scientific evidence to support expert testimony. This decision will surely prevent some mistakes that might have occurred in the past that resulted from overly simple admissibility rules which may have favored plaintiffs. However, as we have argued above, some courts and commentators risk erring in the other direction by instituting or recommending overly simple and excessively stringent admissibility rules. We have argued for a more moderate position between these extremes which would recognize all the evidence on which scientists rely; give due weight to avoiding both legal false positives and legal false negatives for judging the accuracy of tort law decisions; take into account the state of scientific information at the time of trial for judging whether justice has been served; try to preserve procedural fairness between plaintiffs and defendants; and preserve the present balance *77 of interests between adversaries. Adoption of these recommendations will not make judges' jobs easier; bright line rules would do that. On the contrary, the recommendations will make judges' roles more difficult because they will require a more sensitive evaluation and weighing of evidence as envisioned by the Court in Daubert. Taken together, however, these recommendations will make tort law more accurate and fair to both parties.

Footnotes

^a Professor of Philosophy, University of California, Riverside; B.A. Mathematics, University of Colorado (1966); Ph.D., UCLA (1971); M.S.L., Yale Law School (1981).

^{aa} Ph.D. student, Department of Philosophy, University of California, Riverside; B.A., Wesleyan University (1985); J.D., University of California, Davis (1991).

^{aaa} Associate Professor, Environmental Toxicology, University of California, Riverside; B.S., Brigham Young University (1980); M.S., Brigham Young University (1983); Ph. D., University of California, Berkeley (1987); Postdoctoral Fellow, Lawrence Livermore National Laboratory (1987-1989).

Authors' Note: Valuable discussion of and suggestions for this paper were provided by the UCLA Law and Philosophy Discussion Group (which includes Steve Munzer, Herbert Morris, Marshall Cohen, David Dolinko, Barbara Herman, Craig Ihara, Peter Aranella, and Andrews Reath) as well as by Joe Cecil of the Federal Judicial Center, Vernon Walker of the Hofstra University Law School, and Paul Hoffman of the UCR Philosophy Department. We also wish to acknowledge the research assistance provided by Christopher Alexander, a University of Southern California J.D. candidate.

¹ 509 U.S. 579 (1993).

² See *Frye v. U.S.*, 293 F. 1013 (D.C. Cir. 1923) (holding that novel scientific evidence of methodology had to have "general acceptance" in the relevant scientific community to be admitted for consideration at criminal trial).

³ See *Daubert*, 509 U.S. at 587.

⁴ *Id.* at 590.

⁵ *Id.* at 589. The dissenting opinion concurred in rejecting the Frye rule and also concurred that scientific testimony must be relevant, but argued that the majority had gone too far in arguing for the "reliability" of evidence as part of *Rule 702*. See *id.* at 599-600 (Rehnquist, C.J., dissenting).

⁶ *Id.* at 589.

⁷ *Id.* at 590-91.

⁸ *Id.* at 592-93.

9 See [id.](#) at 589-97.

10 What may have misled some judges and some commentators is a failure to distinguish between considerations which bear on the strength of an evidentiary claim with considerations of whether such evidence is relevant to a causal explanation. See [infra](#) notes 73-75 and accompanying text.

11 See [Daubert](#), 509 U.S. 579, 582 (1993).

12 See [id.](#)

13 See [id.](#)

14 See [id.](#)

15 See [id.](#) at 583.

16 See [id.](#) at 583-84.

17 See [Daubert v. Merrell Dow Pharm., Inc.](#), 727 F. Supp. 570, 575 (S.D. Cal. 1989), aff'd, 951 F.2d 1128 (9th Cir. 1991), vacated, 509 U.S. 579 (1993).

18 See [id.](#) at 572.

19 See [Daubert v. Merrell Dow Pharm., Inc.](#), 951 F.2d 1128, 1131 (9th Cir. 1991), vacated, 509 U.S. 579 (1993).

20 See [id.](#) at 1129-30.

21 See [id.](#) at 1130.

22 See [id.](#) The existence of a large amount of contrary evidence seems relevant to the issue of sufficiency of the evidence, not the issue of admissibility. We discuss this point in detail below, as many lower courts appear to have confused the two issues. See [infra](#) notes 73-75 and accompanying text.

23 See [Daubert](#), 951 F.2d at 1131.

24 See [Daubert](#), 509 U.S. at 597-98.

25 See [id.](#) at 589.

26 [Id.](#) at 588 (quoting [Fed. R. Evid. 702](#)).

27 [Id.](#)

28 See [id.](#) at 598 (Rehnquist, C.J., dissenting).

29 See [id.](#) at 590.

30 See [id.](#) at 590-91.

31 See [id.](#)

32 See [id.](#) at 593-94.

33 See [id.](#)

34 [Id.](#) at 596.

35 See [id.](#)

36 [Id.](#) at 596-97.

37 See id.

38 See id. at 597.

39 See id.

40 See id. at 597.

41 For a useful collection of works on the implications of Daubert for toxic tort, pharmaceutical, and products liability cases, see [15 Cardozo L. Rev. 1745-2294 \(1994\)](#).

42 See David E. Bernstein & Peter Huber, Defense Perspective, 1 Shephard's Expert & Sci. Evidence Q. 59 (1993); Marc Whithead, Daubert Will Allow More Expert Testimony, Complicate Jurors' Job, Prejudice Defense, 21 Prod. Safety & Liab. Rep. (BNA) 41 (Summer-Fall 1993).

43 See, e.g., Michael D. Green, Relief at the Frying of Frye: Reflections on *Daubert v. Merrell Dow Pharmaceuticals*, 1 Shephard's Expert & Sci. Evidence Q. 43, 47-48 (1993).

44 See, for example, changes in the National Toxicology Program's criteria for classifying substances as "reasonably anticipated to be human carcinogens," which incorporate considerations, as we argue later in Part IV.B, that might well be excluded as inadmissible in toxic tort suits according to some decisions. See National Toxicology Program, National Institute of Environmental Health Sciences, Review of the Criteria and the Process for Preparing the Biennial Report on Carcinogens Completed: Changes Effective Immediately, in National Toxicology Program Update 3 (1996).

45 Reference Manual on Scientific Evidence (Fed. Judicial Ctr. ed., 1994) [hereinafter Reference Manual]. It should be noted that the essays in this anthology do not necessarily express common guiding principles for the admissibility of evidence, and thus leave a number of issues unaddressed. See Joe S. Cecil, [Reference Manual on Scientific Evidence: Limitations and Potential](#), [36 Jurimetrics J. 225, 225 \(1996\)](#) (noting that the Reference Manual stops "short of advising judges on how to rule regarding difficult issues presented by scientific testimony").

46 Aside from the authority of Daubert, another argument in favor of relatively stringent admissibility standards to combat the proliferation of "junk science" in the courtroom goes as follows: Scientific experts and scientific evidence generally have, for the jury, an aura of authoritativeness and reliability, regardless of actual scientific merit. Thus, juries might assign too much weight to testimony that scientists would consider scientifically unreliable. Therefore, the judge should exclude such evidence, unless it is clearly scientifically reliable. In short, this argument expresses a doubt that juries can appropriately distinguish between "good" scientific testimony and "bad" scientific testimony.

For the sake of argument, we could assume that juries may have this problem (although that is not clear). One difficulty with the above argument becomes apparent when one reflects upon other kinds of evidence which courts routinely admit and in fact make very difficult to impeach. The best example is eyewitness testimony. It is hardly necessary to make the point that courts routinely admit eyewitness testimony if the proffered witness was in fact a percipient witness and the proffered testimony is relevant (assuming various other limitations on admissibility, such as hearsay, prejudice, etc. do not apply). A substantial body of evidence exists which casts severe doubt on the reliability of eyewitness testimony generally, as well as in cases where, for example, environmental factors such as light, noise and the like, impair the witness' abilities. See, e.g., Brian L. Cutler & Steven D. Penrod, *Mistaken Identification* 13 (1995) (citing studies which show an average of 35 percent error in eyewitness identifications). It is clear that juries assign a tremendous amount of weight to such testimony, such that an eyewitness identification can sometimes solely determine the outcome of a case. Furthermore, it is not clear that the conventional safeguards of cross-examination and jury instructions can mitigate the effect of such testimony. Cutler & Penrod, *supra*, at 168, 209 (noting that cross-examination does not provide strong safeguard against the influence of unreliable eyewitness testimony). Nor do jury instructions provide a reliable constraint. See *id.* at 263-264. Cutler and Penrod conclude that expert testimony about the reliability of eyewitness testimony provides the strongest safeguard, but that courts frequently refuse to allow such testimony. See *id.* at 51-52, 251. Thus, testimony which scientists consider very unreliable is frequently admitted for the jury to assign appropriate weight. Why should scientific evidence be treated differently from such testimony?

47 See [Marc S. Klein, After Daubert: Going Forward With Lessons From The Past](#), [15 Cardozo L. Rev. 2219, 2220 \(1994\)](#).

48 See *id.* at 2219.

49 See id. at 2222 (citing [Wells v. Ortho Pharmaceutical Corp.](#), 615 F. Supp. 262 (N.D. Ga. 1985), aff'd in part, and rev'd in part, 788 F.2d 471 (11th Cir.), cert. denied, [479 U.S. 950](#) (1986)).

50 See id. at 2234-35. It may be difficult to place the Daubert decision on an overall liberal-conservative spectrum. On the one hand, the opinion appears to admit a wider range of novel, reasonably-founded scientific evidence than the Frye rule might have, but, on the other hand, it may result in the exclusion of previously admissible forensic evidence such as hand-writing analysis.

51 See [Daubert](#), 509 U.S. at 592.

52 Klein, *supra* note 47, at 2220.

53 As the discussion below suggests, there is a wide range of legitimate scientific evidence that would be excluded by some of the restrictive admissibility rules proposed by the Defense bar. See *infra* Part III.B.

54 See [Daubert](#), 509 U.S. at 597.

55 [57 F.3d 428](#) (5th Cir. 1995).

56 The court stated:
To iterate, we do not now hold that polygraph examinations are scientifically valid or that they will always assist the trier of fact, in this or any other individual case. We merely remove the obstacle of the *per se* rule against admissibility, which was based on antiquated concepts about the technical ability of the polygraph and legal precepts that have been expressly overruled by the Supreme Court.
Assuming that polygraph evidence satisfies the requirements of [Rule 702](#) does not end the inquiry. Other evidentiary rules, such as Rule 403, may still operate to exclude the evidence. While not discussed at length in *Daubert*, the presumption in favor of admissibility established by rules 401 and 402, together with *Daubert*'s "flexible" approach, may well mandate an enhanced role for Rule 403 in the context of the *Daubert* analysis, particularly when the scientific or technical knowledge proffered is novel or controversial.
Id. at 434-35 (citation omitted).

57 [35 F.3d 717](#) (3d Cir. 1994).

58 See *id.* at 742 n.7.

59 See [Daubert](#), 509 U.S. at 597.

60 This observation about the *Daubert* opinion was originally discussed in David L. Faigman et al., [Check Your Crystal Ball At The Courthouse Door, Please: Exploring The Past, Understanding The Present, And Worrying About The Future Of Scientific Evidence](#), 15 Cardozo L. Rev. 1799, 1817-20 (1994). This article utilizes some of their insights. Rule 104(a) requires preliminary questions of qualification of a witness or admissibility of evidence to be determined by the court subject to 104(b). See [Fed. R. Evid. 104\(a\)](#). Rule 104(b) states that relevancy of evidence is conditioned on fact and shall be admitted subject to submission of evidence sufficient to support a finding of fulfillment of that condition. See [Fed. R. Evid. 104\(b\)](#).

61 See Faigman et al., *supra* note 60, at 1817.

62 See *id.* at 1819.

63 See *id.* at 1818-19.

64 See *id.* at 1817-18.

65 See [Daubert](#), 509 U.S. at 592-95.

66 The methodology/conclusions distinction is even more problematic, since in places the Court refers to "reasoning or methodology." See *id.* at 593.

67 [874 F. Supp. 1441](#) (D.V.I. 1994).

68 See *id.* at 1453-55.

69 See *id.* at 1482.

70 See *id.* at 1477, 1483.

71 See *In re Paoli R.R. Yard PCB Litigation*, 35 F.3d 717, 743 n.9 (3d Cir. 1994), cert. denied, 115 S. Ct. 1253 (1995). Later, the Paoli court stated:

The methodology/conclusion distinction remains of some import, however, to the extent that there will be cases in which a party argues that an expert's testimony is unreliable because the conclusions of an expert's study are different from those of other experts. In such cases, there is no basis for holding the expert's testimony inadmissible.

Id. at 746 n.15 (citation omitted) (emphasis added); see also *Cavallo v. Star Enter.*, 892 F. Supp. 756, 769 (E.D. Va. 1995) (excluding plaintiff's scientific evidence even though the methodology used by the expert was acceptable because the conclusions he reached were not a reasonable application of that methodology). The Star Enterprise court relied on some language in Paoli and paid attention to the fact that the defendants had produced several studies which reached a conclusion contrary to that of the plaintiff's expert. See *id.* at 768. Again, the existence of contradictory evidence should not bear on the question of the admissibility of scientific evidence, but rather the sufficiency of the evidence which may be used to avoid a summary judgment.

72 A court, of course, may prevent the jury from hearing the case through, for example, a summary judgment procedure which can, to some degree, involve an evaluation of the strength of the expert's testimony and possibly conclude that no reasonable jury would accept that testimony when compared with testimony offered by an adversary.

73 In Paoli, the court stated in relevant part:

A judge will often think that an expert has good grounds to hold the opinion that he or she does even though the judge thinks that the opinion is incorrect The judge might think that there are good grounds for an expert's conclusion even if the judge thinks that there are better grounds for some alternative conclusion, and even if the judge thinks that a scientist's methodology has some flaws such that if they had been corrected, the scientist would have reached a different result.

35 F.3d at 744.

74 See *Joseph Sanders, Scientific Validity, Admissibility, and Mass Torts after Daubert*, 78 Minn. L. Rev. 1387, 1341 (1994) (arguing that courts in Bendectin cases used inadmissibility to control juries from rendering verdicts to severely injured plaintiffs). He points out that the 40 percent success rate of Bendectin plaintiffs whose cases reached the jury mirrors that of products liability cases in general. See *id.* at 1433. According to Sanders, however, judges remain skeptical of a jury's ability to arrive at correct decisions in cases involving extremely complicated scientific evidence, as is frequently the case in toxic tort actions. See *id.* at 1431-32. Sanders argues that this method of ad hoc jury control is flawed in that courts are dealing with questions of sufficiency using the admissibility rules. See *id.* at 1433. Thus, these cases are driving the development of evidence law to deal with sufficiency questions.

75 Judge Kozinski, writing the opinion in the Daubert remand, affirmed the trial court decision by focusing on the relevancy of the testimony of plaintiffs' experts. Those experts could not testify that Bendectin more than doubled the likelihood of the presence of birth defects but only that Bendectin was capable of causing defects and might only increase the rate of birth defects by 60 percent, not by the 100 percent the Court demanded. Therefore, their testimony was not helpful to the trier of fact and inadmissible under the second prong of *Rule 702*. See *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1321 (9th Cir.), cert. denied, 116 S. Ct. 189 (1995). In addition, in dicta, Kozinski added a factor to the Daubert analysis: namely, whether the expert is testifying based on research she has conducted independently of the litigation, rather than expressly for the purposes of the litigation. Kozinski concludes that if the testimony is based on research conducted independently of the pending litigation, that fact provides "important, objective proof that the research comports with the dictates of good science." *Id.* at 1317. If this is not the case, according to the opinion, the party proffering the testimony must come up with other objective proof of the scientific validity of the technique. See *id.* at 1317-18.

This new factor appears to confuse the Daubert analysis. First, many scientific techniques would not be developed except within the context of litigation, or for regulatory purposes. The reason is that many tort law and regulatory cases concern particular products, particular substances in products, or pollutants of particular processes. These are not typically objects of general scientific investigation unless a firm seeks to develop a product or a process and unless an adversary thinks that the product or process has harmed someone. The vast majority of scientists do not typically investigate such questions. In

many cases, such substances or products would only be the object of inquiry if they became an issue for regulation or tort law. See, e.g., *Star Enterprise*, 892 F. Supp. at 769 (regarding a question of the toxicity of a particular commercial brand of jet fuel). It is unlikely that any research scientist has studied the toxicity of this substance. Second, if an individual has conducted research independently of the litigation, that does not prove that the research has resulted in, or was based upon, a scientifically valid technique. Suppose a litigant retains an expert simply because that expert's flawed research is against the vast weight of research in the particular field. In that case, the research would have no more scientific validity than research conducted specifically for the purposes of litigation. Regardless of whether the research was conducted independently or for the purposes of litigation, it should be judged on its own merits. See *Valentine v. Pioneer Chlor Alkali Co., Inc.*, 921 F. Supp. 666 (D. Nev. 1996), in which the court followed the Daubert remand court in focusing upon whether the proffered expert relied upon research conducted independently of litigation. Similar arguments support the claim that there should not be too much reliance on peer-reviewed evidence as well.

76 See, e.g., *American & Foreign Ins. Co. v. General Elec. Co.*, 45 F.3d 135, 139 (6th Cir. 1995) (lower court exclusion of expert testimony on circuit breaker design must be clearly erroneous to show abuse of discretion); *United States v. Dorsey*, 45 F.3d 809, 815-16 (4th Cir. 1995) (applying abuse of discretion standard to lower court ruling on admissibility of forensic anthropologist's testimony); *Bradley v. Brown*, 42 F.3d 434, 436-37 (7th Cir. 1994) (holding that lower court's findings regarding doctors' testimony will not be overturned "unless they are manifestly erroneous").

77 See *Cook v. American S.S. Co.*, 53 F.3d 733, 738 (6th Cir. 1995).

78 See *Paoli*, 35 F.3d at 741-52.

79 See *In re Paoli R.R. Yard PCB Litigation*, 706 F. Supp. 358, 361 (E.D. Pa. 1988), rev'd, 916 F.2d 829 (3d Cir. 1990), cert. denied, 499 U.S. 961 (1991).

80 See *id.* at 375.

81 These factors were derived from *United States v. Downing*, 753 F.2d 1224, 1238-39 (3d Cir. 1985). See *Paoli*, 35 F.3d at 742.

82 *Paoli*, 35 F.3d at 742.

83 In Daubert, the Court highlighted the "liberal thrust" of the Federal Rules' increasingly relaxed standards of admissibility of opinion testimony, and denounced the stringency of the Frye test. The Court held that scientific evidence would be admissible if the underlying reasoning and methodology were valid. The Court explained that "many factors will bear on the inquiry, and we do not presume to set out a checklist." *Daubert*, 509 U.S. at 592. A district court, in TMI Litigation Cases Consolidated II, appears to add some fairly stringent considerations to those of Paoli and Daubert. See 922 F. Supp. 1038, 1038-1046 (D. Pa. 1996) (giving special consideration to testable hypotheses, peer review, potential rate of error, existence of standard controlling techniques used, general acceptance of methodology, relationship between method and techniques, qualifications of expert based on methodology, and nonjudicial uses of the methodology).

84 See *Paoli*, 35 F.3d at 743.

85 The court stated:

Thus, as we explained above, we think that the primary limitation on the judge's admissibility determinations is that the judge should not exclude evidence simply because he or she thinks that there is a flaw in the expert's investigative process which renders the expert's conclusions incorrect. The judge should only exclude the evidence if the flaw is large enough that the expert lacks "good grounds" for his or her conclusions.

Id. at 746.

86 See *id.* at 745.

87 See *id.* at 750.

88 See *id.* (quoting *Brody v. Spang*, 957 F.2d 1108, 1115 (3d Cir. 1992)).

89 See *id.* at 749.

90 See supra Parts II.C.1 and 2.

91 See [Rubanick v. Witco Chem. Corp.](#), 593 A.2d 733, 747-49 (N.J. 1991) (conceding that tort law cannot demand the same high level of proof for theories of causation that is required by scientific method). The court found that scientific theories of causation that are reliable, reasonable, and “proffered by an expert who is sufficiently qualified” would be admissible. *Id.* See also [Landrigan v. Celotex Corp.](#), 605 A.2d 1079, 1084 (N.J. 1991) (adopting a “broad[er] standard for determining the reliability and admissibility of scientific theories of causation in toxic-tort litigation” than set forth in *Rubanick*). The *Landrigan* court’s standard based admissibility of expert opinion on “the validity of the expert’s reasoning and methodology.” *Id.* at 1084. This point was suggested to the authors by Vern Walker.

92 Cecil, supra note 45, at 229.

93 Unlike criminal law which embodies procedural rules that tend to prevent the wrongful conviction of innocent persons, tort law more equally balances the concerns of avoiding wrongly holding defendants accountable and wrongly exonerating them. There is substantial legal history supporting this view. In [Speiser v. Randall](#), 357 U.S. 513 (1958), Justice Brennan noted: There is always in litigation a margin of error, representing error in factfinding, which both parties must take into account. Where one party has at stake an interest of transcending value--as a criminal defendant his liberty--this margin of error is reduced as to him by the process of placing on the other party the burden of ... persuading the factfinder at the conclusion of the trial of his guilt beyond a reasonable doubt.

Id. at 525-26.

This theme was also developed by Justice Harlan in a concurring opinion in [In re Winship](#), 397 U.S. 358 (1970):

The standard of proof influences the relative frequency of these two types of erroneous outcomes. If, for example, the standard of proof for a criminal trial were a preponderance of the evidence rather than proof beyond a reasonable doubt, there would be a smaller risk of factual errors that result in freeing guilty persons, but a far greater risk of factual errors that result in convicting the innocent. Because the standard of proof affects the comparative frequency of these two types of erroneous outcomes, the choice of the standard to be applied in a particular kind of litigation should, in a rational world, reflect an assessment of the comparative social disutility of each.

Id. at 371. Note that the two types of erroneous outcomes possible are a factual outcome which favors the plaintiff when the facts warrant an outcome for the defendant or an “erroneous factual determination” for the defendant when a correct understanding justifies a judgment for the plaintiff. Justice Harlan then discussed the preponderance of the evidence standard: In a civil suit between two private parties for money damages, for example, we view it as no more serious in general for there to be an erroneous verdict in the defendant’s favor than for there to be an erroneous verdict in the plaintiff’s favor. A preponderance of the evidence standard therefore seems peculiarly appropriate for, as explained most sensibly, it simply requires the trier of fact to “believe that the existence of a fact is more probable than its nonexistence”

Id. at 371-72.

The Supreme court has also referred to the notion of “comparative social disutility” in more recent cases. See, e.g., [Santosky v. Kramer](#), 455 U.S. 745, 755 (1982) (adopting the standard set forth in *Addington* that “in any given proceeding, the ... standard of proof ... reflects not only the weight of the private and the public interests affected, but also a societal judgment about how the risk of error should be distributed between the litigants”); [Addington v. Texas](#), 441 U.S. 418, 423 (1979) (explaining that the preponderance of the evidence requires litigants to “share the risk of error in roughly equal fashion”).

94 This distinction is a concern in science for a number of reasons. It helps ensure that a scientist’s enthusiasm for his or her own work does not inadvertently overwhelm his or her impartial judgment. Scientists want to be sure that contributions to the stock of knowledge are well-justified and not the result of overzealous advocacy, random chance, or other factors which might lead to mistaken additions to the body of scientific knowledge. It also prevents scientific research from chasing elusive chimera. See generally, [Michael Green, Expert Witnesses and Sufficiency of Evidence](#), 86 Nw. U. L. Rev. 643, 687-89 (1992) (discussing the inverse relationship between false negatives and false positives).

95 This appears to be the thrust of the Supreme Court’s comments in the cases quoted supra note 93.

96 See John Rawls, *A Theory of Justice* 85 (1971).

97 The court in [Wade-Greux v. Whitehall Laboratories, Inc.](#), 874 F. Supp. 1441 (D.V.I.), aff’d, 46 F.3d 1120 (3d Cir. 1994), appears to require such stringent evidentiary standards that toxicologists would have to be virtually certain that a substance is a teratogen before that conclusion could be admitted for tort law purposes. The more courts or commentators insist on

certainty before scientific conclusions are permitted, the more such a standard approaches the second notion of imperfect procedural justice. Such a standard, by requiring the near certainty of conclusions based on it, more nearly assures scientists that it would stand the test of time. Moreover, such a standard approximates one which compares the evidence in a trial with the scientific conclusion that would be arrived at once all the evidence was submitted. For the view of one toxicologist who appears to demand virtual certainty before declaring a substance a human carcinogen, see generally Arthur Furst, Yes, But is it a Human Carcinogen?, 9 J. Am. C. Toxicology 1 (1990).

98 See [C.P. Gillette & J.E. Krier, Risk, Courts and Agencies](#), 138 U. Pa. L. Rev. 1027, 1043-61 (1990) (arguing, in the context of addressing concerns about “junk science,” that procedural rules favoring plaintiffs, including causation rules, merely help to balance out a bias against plaintiffs in obtaining access to tort remedies). Gillette and Krier also argue that the overall balance of interests between plaintiffs and defendants is appropriate. See *id.*

99 Two cases that appear to recognize this point are [Rubanick v. Witco Chemical Corp.](#), 593 A.2d 733, 747-49 (N.J. 1991) (conceding that while the scientific method requires theories of causation to be generally accepted by the relevant scientific community, tort law employs such theories for different purposes, and will thus admit theories of causation that are not widely accepted, but are nevertheless based on sound methodology and reasoning), and [Landrigan v. Celotex Corp.](#), 605 A.2d 1079, 1088-89 (N.J. 1991) (explaining that tort cases involving latent diseases or diseases of unknown origin require an even more lenient standard of admissibility for theories of causation than the one set forth in Rubanick). This point was suggested to the authors by Vern Walker.

100 See *infra* Parts III.B.2-3.

101 See *infra* Parts III.B.2, .B.5.

102 See, e.g., [Wade-Greux](#), 874 F. Supp. at 1478-79 (holding expert teratogenicity testimony inadmissible because methodologies were neither generally accepted within scientific community nor subjected to peer review).

103 See *id.*

104 The insistence on certain demanding tests of statistical significance or other stringent evidentiary measures may be exhibited here. See Carl F. Cranor, Regulating Toxic Substances: A Philosophy of Science and the Law 71-81 (1993).

105 See [Daubert v. Merrell Dow Pharm.](#), 43 F.3d 1311, 1314 (9th Cir.), cert denied, 116 S. Ct. 189 (1995).

106 874 F. Supp. at 1450-51 (highlighting the fact that the relevant scientific community's criteria regarding methodology for determining teratogenicity required positive human epidemiological studies but that the testimony of proffered expert witnesses was not supported by such studies).

107 See [Daubert](#), 509 U.S. at 593-95.

108 [Wade-Greux](#), 874 F. Supp. at 1448.

109 *Id.* at 1450.

110 Summary judgment was granted because the plaintiff's evidence fell far short of the court's articulated criteria, see *id.* at 1476-1486, and the decision was upheld without comment on appeal. See [Wade-Greux v. Whitehall Lab., Inc.](#), 46 F.3d 1120 (3d Cir. 1994) (unpublished table decision). Nonetheless, the announced criteria seem particularly problematic to us.

111 See [Daubert](#), 509 U.S. at 595 (“The focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.”).

112 The court listed various factors for determining whether an agent is a human teratogen, noting that those factors (which included epidemiological evidence, animal models, and a demonstrated dose-response relationship) were generally accepted by the community of teratologists, taught in medical schools throughout the country, and included in highly-esteemed treatises on teratology. See [Wade-Greux](#), 874 F. Supp. at 1450-51.

113 In its findings of fact on Epidemiology, the court stated:

Absent consistent, repeated human epidemiological studies showing a statistically significant increased risk of particular birth defects associated with exposure to a specific agent, the community of teratologists does not conclude that the agent is a human teratogen.

Id. at 1453 (citations omitted). Scientist Arthur Furst appears to adopt a similar view, see *Furst*, *supra* note 97, at 12, but at least it is excusable for the context within which he works. He was addressing a society of toxicologists and asking what criteria a scientist should require to be satisfied before being certain on substantive scientific grounds that a substance is a carcinogen. We believe that courts assessing the admissibility of scientific evidence are operating in a much different context with different evidentiary rules and different guidance from the *Daubert* decision.

114 See *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 589 (1993).

115 See *infra* notes 153-64 and accompanying text.

116 See *Daubert*, 509 U.S. at 596.

117 See *Wade-Greux*, 874 F. Supp. at 1452.

118 See *infra* notes 137-43 and accompanying text.

119 See *infra* notes 122-26 and accompanying text.

120 *In re Agent Orange Prod. Liab. Litig.*, 611 F. Supp. 1223 (E.D.N.Y. 1985).

121 *Id.* at 1231. While this statement is ambiguous as to whether it is a claim about the particular case or a more general criterion for admissibility, a number of courts appear to have taken his remarks as announcing a general criterion. See cases cited *infra* note 125.

122 *Lynch v. Merrell-National Lab., Div. of Richardson-Merrell, Inc.*, 830 F.2d 1190 (1st Cir. 1987).

123 *Id.* at 1194.

124 Green, *supra* note 94, at 665 n.101.

125 See *Brock v. Merrell Dow Pharm., Inc.*, 874 F.2d 307, 312 (5th Cir. 1989) (deciding the case on sufficiency of evidence reasons, the court concluded that a Bendectin plaintiff must proffer a statistically significant study before satisfying her burden of proof on causation), cert. denied, 494 U.S. 1046 (1990); *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823, 825, 831 n.59 (D.C. Cir. 1988) (noting that “epidemiological studies are of crucial significance”), cert. denied, 493 U.S. 882 (1989).

126 See *Renaud v. Martin Marietta Corp.*, 749 F. Supp. 1545 (D. Colo. 1990); *Carroll v. Litton Sys., Inc.*, No. B-C-88-253, 1990 WL 312969, at *47 (W.D.N.C. Oct. 29, 1990); *Thomas v. Hoffman-La-Roche, Inc.*, 731 F. Supp. 224 (N.D. Miss. 1989), aff'd on other grounds, 949 F.2d 806 (5th Cir. 1992).

127 See, e.g., *Marder v. G.D. Searle & Co.*, 630 F. Supp. 1087 (D. Md. 1986).

128 A strict definition of “epidemiological evidence” should include case reports and clinical studies, for these are instances of human evidence. Courts frequently exclude case reports out of hand. See, e.g., *Cavallo v. Star Enter.*, 892 F. Supp. 756, 765 (E.D. Va. 1995) (dismissing a peer-reviewed article because it “is an anecdotal case report and does not reflect the results of pre-designed study in a controlled setting”). This is a mistake in our judgment, although we do not pursue it further. Case reports with the proper background conditions, for example, of a particularly rare disease, may be quite good evidence.

129 See *infra* Part IV.B.

130 See *infra* Parts III.B.4.a.-d.

131 See, e.g., *Brock v. Merrell Dow Pharm., Inc.*, 874 F.2d 307, 312 (5th Cir. 1989), cert. denied, 494 U.S. 1046 (1990).

132 See, e.g., *Daubert v. Merrell Dow Pharm.*, 43 F.3d 1311, 1321 (9th Cir.), cert. denied, 116 S. Ct. 189 (1995).

133 See, e.g., David E. Bernstein, *The Admissibility of Scientific Evidence After Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 15 *Cardozo L. Rev.* 2139, 2167-70 (1994) (discussing the imposition of nine “aspects” of a statistical association between two variables, which were first proposed by Sir Austin Bradford Hill in 1965); see also *infra* Part III.B.4.d.

134 See, e.g., *Brock*, 874 F.2d at 312; Development in the Law-- *Confronting the New Challenges of Scientific Evidence*, 108 *Harv. L. Rev.* 1532, 1542 (1995).

135 See David Ozenhoff & Leslie I. Boden, *Truth and Consequences: Health Agency Responses to Environmental Health Problems*, 12 *Sci., Tech. & Hum. Values* 70, 73-74 (1987).

136 See *id.* at 74; Tom Christoffel & Stephen P. Teret, *Epidemiology and the Law: Courts and Confidence Intervals*, 81 *Am. J. Pub. Health* 1661, 1665 (1991).

137 See Amicus Brief of Professor Kenneth Rothman et al. In Support of Petitioners at 3-7, *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993) (No. 92-102).

138 See *id.*

139 See *id.*; see also Amicus Brief of Professor Alvan R. Feinstein in Support of Respondent at 4-8, *Daubert v. Merrell-Dow Pharm., Inc.*, 509 U.S. 579 (1993) (No. 92-102).

140 For discussion of this point, see Joseph L. Fleiss, *Significance Tests Have a Role in Epidemiologic Research: Reactions to A.M. Walker*, 76 *Am. J. Pub. Health* 559, 559-60 (1986); Steven N. Goodman & Richard Royall, *Evidence and Scientific Research*, 78 *Am. J. Pub. Health* 1568, 1568-74 (1988); Green, *supra* note 94, at 685; Charles Poole, *Beyond the Confidence Interval*, 77 *Am. J. Pub. Health* 195, 195-99 (1987); W. Douglas Thompson, *Statistical Criteria in the Interpretation of Epidemiologic Data*, 77 *Am. J. Pub. Health* 191, 191-94 (1987); Alexander M. Walker, *Reporting the Results of Epidemiologic Studies*, 76 *Am. J. Pub. Health* 556, 556-58 (1986).

141 Green, *supra* note 94, notes one reviewer who identified seventy-one epidemiologic studies that failed to satisfy statistical significance, but concluded that the “studies were consistent with a moderate or strong effect of the treatment under investigation.” *Id.* at 685 (citing Jennie A. Freiman et. al., *The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial: Survey of 71 “Negative” Trials*, 299 *New Eng. J. Med.* 690 (1978)).

142 See *id.* at 686.

143 See *id.*; see also Cranor, *supra* note 104, at 71-81.

144 See Green, *supra* note 94, at 681.

145 See *id.* at 682.

146 See *id.* at 683.

147 See *id.* at 687; Cranor, *supra* note 104, at 30-39.

148 See Green, *supra* note 94, at 683 (“There is an inverse relationship between these two types of errors in any given study; as one is reduced the other is increased.”).

149 See Cranor, *supra* note 104, at 36-40.

150 See Green, *supra* note 94, at 691-92 (providing an example showing that the chances of a false negative can easily be nearly 10 times the chances of a false positive); see also Cranor, *supra* note 104, at 71-78.

151 See Green, *supra* note 94, at 691-692; Cranor, *supra* note 104, at 71-78.

152 See Green, *supra* note 94, at 693.

153 Austin Bradford Hill, The Environment and Disease: Association or Causation?, 58 Proceedings of the Royal Society of Medicine 295, 299 (1965), reprinted in Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods 15, 19 (Sander Greenland ed., 1987).

154 Green, *supra* note 94, at 693-94.

155 See, e.g., *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1320-21 (9th Cir. 1995), on remand from 509 U.S. 579 (1993); see also *In re Joint E. & S. Dist. Asbestos Litig.*, 758 F. Supp. 199, 203 (S.D.N.Y. 1991), aff'd, 52 F.3d 1124 (2d Cir. 1995). Commentators who argued for this early on are Bert Black and David Lilienfeld. See *Bert Black & David E. Lilienfeld, Epidemiologic Proof in Toxic Tort Litigation*, 52 Fordham L. Rev. 732, 769 (1984).

156 See *Daubert*, 43 F.3d at 1320-22.

157 See *id.* at 1320-21.

158 See *id.* at 1321.

159 See *id.* at 1322.

160 Green, *supra* note 94, at 691.

161 Some courts have recognized that epidemiological studies showing a relative risk of less than two might be sufficient to show tort law causation where other risk factors can be eliminated. See *Gassis v. Johns-Manville Corp.*, 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991) (ruling that testimony concerning epidemiological studies showing a relative risk less than 2.0 is admissible as one basis for an expert's opinion). But cf. *In re Joint E. & S. Dist. Asbestos Litig.*, 758 F. Supp. 199, 203 (S.D.N.Y. 1991) (ruling that for epidemiological evidence to establish causation by a preponderance of the evidence, the risk level proven by plaintiff must exceed 2.0 in absence direct evidence of causation), aff'd, 52 F.3d 1124 (2d Cir. 1995).

162 See H. Kato, Cancer Mortality, in *Cancer in Atomic Bomb Survivors* (I. Shigematsu & A. Kagan eds., Japan Scientific Societies Press 1986), quoted in Arthur K. Sullivan, *Classification, Pathogenesis, and Etiology of Neoplastic Diseases of the Hematopoietic System*, in G.R. Lee et al., 2 *Wintrobe's Clinical Hematology* 1725, 1750 (Lea & Febiger eds., 9th ed. 1993).

163 See *Julius C. McElveen & Chris Amantea, Risk Symposium: Legislating Risk Assessment*, 63 U. Cin. L. Rev. 1553, 1556 (1995).

164 See, e.g., Frederica Perera, *Molecular Epidemiology: Insights into Cancer Susceptibility, Risk Assessment, and Prevention*, 88 J. Nat'l Cancer Inst. 496, 496-509 (1996) (discussing the interaction of environmental factors and genetic and acquired susceptibilities to cancer).

165 See W. Page Keeton et al., *Prosser and Keeton on Torts* 291-92 (1984).

166 See *supra* Parts III.B.4.a.-b for discussion of these two restrictions on epidemiological studies.

167 See Cranor, *supra* note 104, at 36-40.

168 See *id.* at 31-39.

169 This appears to have been a problem with some of the early studies on breast implants. See Shanna H. Swan, *Epidemiology in the Courtroom: The Case of Silicone Breast Implants*, in *Litigating Breast Implant Cases: A Satellite Program*, at 401, 407 (PLI Litig. & Admin. Practice Course Handbook Series No. H4-5149 1992).

170 See Cranor, *supra* note 104, at 34-38.

171 Manolis Kogevinas & Paolo Boffetta, 48 Brit. J. Indus. Med. 575-76 (1991) (letter to editor criticizing a study by O. Wong, *A Cohort Mortality Study and a Case Control Study of Workers Potentially Exposed to Styrene in Reinforced Plastics and Composite Industry*, 47 Brit. J. Indus. Med. 753-62 (1990), for having too short a follow-up--seven years--even though the author had a large sample population to study). The authors claimed that this defect "should caution against a premature negative evaluation of cancer risk in the reinforced plastics industry." *Id.* at 575.

172 See D. Schottenfeld & J.F. Haas, Carcinogens in the Workplace, 29 CA-Cancer J. for Clinicians 144, 156-59 (1979).

173 A National Institute of Occupational Health epidemiological study was noted to have an 80 percent chance of detecting a nine-fold relative risk of bladder cancer in a cohort of workers exposed to 4,4'-methylenebis (2-chloroaniline) (MBOCA) in 1985, but by 1995 the same study would have an 80 percent chance of detecting a four-fold relative risk. See Elizabeth Ward et al., 4,4'-Methylenebis (2-Chloroaniline): An Unregulated Carcinogen, 12 Am. J. Indus. Med. 537, 542 (1987).

174 See Hill, *supra* note 153, at 15-19.

175 Hill's considerations also include the so-called "Koch's Postulates" proposed about 100 years ago for infectious diseases. See Linda A. Bailey et al., Reference Guide on Epidemiology, in Reference Manual, *supra* note 45, at 121, 161. All of Koch's Postulates are included in Hill's considerations except "consideration of alternative explanations," which is always considered for epidemiological studies, usually under the consideration of "confounders." *Id.* at 160-63. Hill's considerations include two additional features not explicitly included in Koch's Postulates: the possibility of appealing to experimental evidence and argument by analogy. See Hill, *supra* note 153, at 18-19.

176 An early work which appears to insist on using Hill's criteria is Black and Lilienfeld, *supra* note 155, at 764 (arguing that all of Koch's hypothesis should be satisfied). See also Bernstein, *supra* note 133, at 2166, 2168. Bernstein's remarks are ambiguous between the claim that all of Hill's criteria must be met for a study to be admissible (a view that is clearly at odds with good scientific practice) and the claim that if none (which would include the temporal criteria) of the criteria are met a study is not admissible. See *id.* at 2168. The second contention we would clearly agree with while the first we sharply disagree with, as did Hill himself. See Hill, *supra* note 153, at 19. Furthermore, Bernstein argues that if "proffered epidemiological evidence meets some but not all of the criteria a judge would do well to consult with a court-appointed epidemiological expert to assist her in judging the reliability of the evidence." Bernstein, *supra*, at 2168-69. Based on Hill and contemporary epidemiologists' views, judges should be hesitant to rule epidemiological studies inadmissible on such grounds since the absence of any single criteria (with the exception of temporality) is consistent with causation.

177 Hill, *supra* note 153, at 16.

178 *Id.*

179 *Id.*

180 *Id.* at 17.

181 See *id.* at 17-18.

182 See *id.* at 17.

183 *Id.*

184 *Id.* at 17.

185 See *infra* notes 198-201 and accompanying text.

186 See Hill, *supra* note 153, at 18.

187 *Id.*

188 See *id.*

189 *Id.*

190 See *id.*

191 *Id.*

192 *Id.*

193 Subsequent studies have provided some evidence for the tumorigenicity of arsenic in animals. See International Agency for Research on Cancer, Arsenic and Arsenic Compounds, in IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Supp. VII 100, 102-03 (1987).

194 See Hill, *supra* note 153, at 18.

195 *Id.* at 19.

196 *Id.*

197 *Id.*

198 Bernstein, *supra* note 133, at 2168 (emphasis added). **Rule 702** allows a qualified expert witness to testify as to scientific, technical or other specialized knowledge that may assist the trier of fact to better understand the evidence or a fact at issue. **Fed. R. Evid. 702**.

199 See Kenneth J. Rothman, Causes, 104 Am. J. of Epidemiology 587, 588 (1976), reprinted in Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods, *supra* note 153, at 40, 41 (noting that sufficient cause may exist even in the absence of individual component causes).

200 See Hill, *supra* note 153, at 15-19.

201 Sander Greenland, Preface to Hill, *supra* note 153, at 14 (citation omitted).

202 *In re Joint E. & S. Dist. Asbestos Litig.*, 827 F. Supp. 1014 (S.D.N.Y. 1993).

203 See *id.* at 1038. The court stated:

While none of the Sufficiency Criteria is decisive by itself in determining the sufficiency of a plaintiff's epidemiological evidence in the context of a Rule 50(b) motion, sufficient epidemiological evidence will necessarily satisfy several of these criteria. More significantly, when epidemiological evidence fails to satisfy any of the Sufficiency Criteria, it cannot be relied on to support a jury verdict in the face of a motion for judgment as a matter of law.

Id.

204 See Hill, *supra* note 153, at 17.

205 See *id.*

206 See *id.* at 18.

207 The *In re Joint Eastern* court noted that "sufficient epidemiological evidence will necessarily satisfy several of these criteria." *In re Joint Eastern*, 827 F. Supp. at 1038. Epidemiologists and the Reference Manual have rejected that view. See *infra* notes 208-16 and accompanying text; see also Bailey et al., *supra* note 175, at 161-64.

208 See Rothman, *supra* note 199, at 43. This is a subtle point concerning a common model of causation. A causal factor for Rothman is a necessary condition of a set of conditions which are jointly sufficient for producing a particular outcome. Rothman added:

A component cause which requires, to complete the sufficient cause, other components with low prevalence is thereby a "weak" (component) cause. The presence of such a component cause modifies the probability of the outcome only slightly, from zero to an average value just slightly greater than zero, reflecting the rarity of the complementary component causes. On the other hand, a component cause which requires, to complete the sufficient cause, other components which are nearly ubiquitous is a "strong" (component) cause. In epidemiologic terms, a weak cause confers only a small increment in disease risk, whereas a strong cause will increase disease risk substantially.

Thus the strength of a causal risk factor depends on the prevalence of the complementary component causes in the same sufficient cause. But this prevalence is often a matter of custom, circumstance or chance, and is not a scientifically generalizable characteristic.... Thus, the strength of a causal risk factor, as it might be measured by the "risk ratio" (relative risk) parameter, is dependent on the distribution in the population of the other causal factors in the same sufficient cause. The term strength of

a causal risk factor retains some meaning as a description of the public health importance of a factor. However, the common epidemiologic parlance about strength of causal risk factors is devoid of meaning in the biologic description of disease etiology. *Id.* at 42-43.

209 Mervyn Susser, Judgment and Causal Inferences Criteria in Epidemiologic Studies, 105 Am. J. of Epidemiology 1, 9 (1977), reprinted in Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods, *supra* note 153, at 69, 77.

210 *Id.*

211 Brian MacMahon & Thomas F. Pugh, Causes and Entities of Disease, in Methods of Preventive Medicine 11, 16 (D.W. Clark & B. MacMahon eds., 1967), reprinted in Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods, *supra* note 153, at 26, 31.

212 Susser, *supra* note 209, at 81.

213 See Bailey et al., *supra* note 175, at 162.

214 *Id.* at 162-63.

215 *Id.* at 163.

216 See Bernstein, *supra* note 133, at 2158-61.

217 *In re Agent Orange Prod. Liab. Litig.*, 611 F. Supp. 1223 (E.D.N.Y. 1985).

218 See, e.g., *In re Paoli R.R. Yard PCB Litig.*, 706 F. Supp. 358, 366-68 (E.D. Pa. 1988), *rev'd*, 916 F.2d 829 (3d Cir. 1990), *cert. denied*, 499 U.S. 961 (1991).

219 *In re Agent Orange*, 611 F. Supp. at 1241.

220 See *id.*

221 *Id.*

222 See, e.g., *Paoli*, 706 F. Supp. at 367-68.

223 *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 779 (3d Cir. 1994) (noting that humans and monkeys are likely to show similar sensitivity to PCBs), *cert. denied*, 115 S. Ct. 1253 (1995).

224 D.P. Rall et al., Alternatives to Using Human Experience in Assessing Health Risks, 8 Ann. Rev. Pub. Health 355, 356 (1987).

225 See *id.*

226 See *id.* The emphasis on "the majority of cases" seems especially germane for the tort law and its ultimate burden of persuasion. If, in the majority of cases, animal responses are similar to human responses, animal evidence should have probative value for judging toxicity to humans in the tort law.

227 *Id.*

228 The International Agency for Research on Cancer is part of the World Health Organization. It is considered by most scientists as the definitive body for the identification of cancer-causing agents in humans. See Bruce N. Ames, The Causes and Prevention of Cancer: The Role of the Environment, in CA51 ALI-ABA 49, 58 (1995).

229 See J. Huff, Chemicals and Cancer in Humans: First Evidence in Experimental Animals, 100 Envtl. Health Persp. 201, 204 (1993); International Agency for Research on Cancer, Preamble, in 63 IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 9, 17 (1995).

230 See Huff, *supra* note 229, at 204.

231 See Filov et al., Quantitative Toxicology 18-21 (1979).

232 See id. Scaling dose-related information from one species to another can vary in complexity from relatively simple to very complicated. For additional information, see id. See also Shayne C. Gad, Model Selection and Scaling, in *Animal Models in Toxicology* 813-40 (Shayne C. Gad & Christopher P. Chengelis eds., 1992).

233 For example, fenclozic acid, a potential anti-inflammatory drug, was seen to cause acute cholestatic jaundice in humans even though studies on a series of other species failed to produce similar toxic effects. See Gad, *supra* note 232, at 818-819. These differences are sometimes widely publicized by special interest groups giving the public the mistaken impression that the results seen in experimental animals have little value for predicting adverse effects in humans. See Shayne C. Chengelis & Christopher P. Gad, *Introduction to Animal Models in Toxicology*, *supra* note 232, at 2.

234 See Filov et al., *supra* note 231, at 18-20; Chengelis and Gad, *supra* note 233, at 1-20.

235 Ping Kwong Chan & A. Wallace Hayes, Principles and Methods for Acute Toxicity and Eye Irritancy, in *Principles and Methods of Toxicology* 169, 206-12 (A. Wallace Hayes ed., 1989); A.T. Mosberg & A. Wallace Hayes, Subchronic Toxicity Testing, in *Principles and Methods of Toxicology*, *supra* at 221, 226-31.

236 See, e.g., Chan & Hayes, *supra* note 235, at 206, 211-12; see also John J. Cohrssen & Vincent T. Covello, *United States Council on Environmental Quality, Risk Analysis: A Guide to Principles and Methods for Analyzing Health and Environmental Risks* 38 (1989).

237 See Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, in *Reference Manual*, *supra* note 45, at 181, 186.

238 See Chan & Hayes, *supra* note 235, at 206, 212.

239 *Id.* at 212.

240 See Rall et al., *supra* note 224, at 356.

241 “These epidemiological studies alone demonstrate that on the basis of present knowledge, there is no question of fact: Agent Orange cannot now be shown to have caused plaintiffs' numerous illnesses.” *In re Agent Orange Prod. Liab. Litig.*, 611 F. Supp. 1223, 1241 (E.D.N.Y. 1985). However, there have been some subsequent studies which show that people exposed to the herbicides used in Agent Orange in occupational or environmental exposure had increased risks of various cancers and other diseases and that at least Vietnam veterans with very slight exposure to Agent Orange “could have risks approaching those in occupational and environmental settings.” Committee to Review the Health Effects in Vietnam Veterans of Exposure to Herbicides, Institute of Medicine, *Veterans and Agent Orange* 14 (1996).

242 See *In re Agent Orange*, 611 F. Supp. at 1241. However, even if defendants had the only good epidemiological studies, it does not follow that animal studies were irrelevant. Plaintiffs' evidence based on animal studies might be considered insufficient in the face of good epidemiological studies on the other side, but as noted above admissibility and sufficiency are questions which arise at different junctures in civil legal procedure.

243 See *supra* Part II.C.2.

244 See, e.g., *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 780 (3d Cir. 1994), cert. denied, 115 S. Ct. 1253 (1995); *Hines v. Consolidated Rail Corp.*, 926 F.2d 262, 271 (3d Cir. 1991); *Villari v. Terminix Int'l, Inc.*, 692 F. Supp. 568, 572 (E.D. Pa. 1988).

245 The Paoli court's own citation provides helpful support:
See, e.g., *In re Bendectin Prod. Liab. Litig.*, 732 F. Supp. 744, 749 (E.D. Mich. 1990) (experts in the field think it is reasonable to rely on non-epidemiological studies to link Bendectin to birth defects); *Hagen v. Richardson-Merrell*, 697 F. Supp. 334, 337 (N.D. Ill. 1988) (Defendants did not adequately demonstrate that expert opinion based partly on animal studies should be excluded); *Saakbo Rubanick v. Witco Chem. Corp.*, 242 N.J. Super. 36, 576 A.2d 4, 7, 15 (1990) (under New Jersey law reversing trial court's exclusion of expert testimony, which was partly based on animal studies that PCBs caused cancer). In *Villari v. Terminix Int. Inc.*, 692 F. Supp. 568, 570 (E.D. Pa. 1988), Judge Pollak explained that:
[W]hile it may be true that defendant can offer tests and experiments that do not support the findings of plaintiff's expert, the defendant cannot deny that animal studies are routinely relied upon by the scientific community in assessing the carcinogenic

effects of chemicals on humans. Even defendant's own expert acknowledges that animal experiment studies are built on 'prudent presumptions,' although he concludes that they should not be admitted.

Paoli, 35 F.3d at 780.

246 Id. (citing *Turpin v. Merrell Dow Pharm., Inc.*, 959 F.2d 1349, 1360 (6th Cir. 1992) (excluding the testimony where the record failed to make clear how the animal studies were sufficient to show that Bendectin causes birth defect more probably than not); *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823, 830 (D.C. Cir. 1988) (excluding animal studies of Bendectin because of the overwhelming body of contrary epidemiological evidence and the admission of the expert that animal studies merely raise a suspicion of causation in humans); *Lynch v. Merrell-Nat. Lab.*, 830 F.2d 1190, 1194 (1st Cir. 1987) (excluding animal studies of Bendectin where they stood in the face of significant contrary epidemiological data); *Viterbo v. Down Chem. Co.*, 826 F.2d 420 (5th Cir. 1987) (excluding the evidence where there was only a single animal study and it showed a link to a disease completely different than plaintiff's diseases); *In re Agent Orange*, 611 F. Supp. at 1241 (excluding animal studies of Agent Orange based partly on the court's earlier conclusion that there was significant epidemiological data, that the Center for Disease Control had concluded that the animal studies did not demonstrate adverse human health effects, and that the animal studies gave pregnant females high doses at critical times)).

247 See Chengelis & Gad, *supra* note 233, at 1-2.

248 For a discussion of the distinction between admissibility and sufficiency review, see *supra* Part II.C.2.

249 See Renate D. Kimbrough, Case Studies, in *Industrial Toxicology* 414, 417-20 (P.L. Williams & J.L. Burson eds., 1985).

250 See *id.* at 420.

251 See *id.* at 419-20.

252 See *id.* at 420.

253 See *id.*

254 See *id.*

255 Kimbrough notes that the evidence that showed the toxicity of dimethylnitrosamine was based on studies in rats and then an amount lethal to adult humans was calculated from the results of those studies. See *id.*

256 See, e.g., *Novak v. United States*, 805 F.2d 718 (6th Cir. 1989) (refusing to allow expert to rely on case studies related to dermatomytosis because causal link was merely conjectural).

257 Indeed, the defendant, who had access to toxins as a researcher responsible for mixing diets for cancer research studies in animals, had had an affair with the wife of one of the decedents. The defendant had confronted the family before with a gun and was, at the time, on parole. See Kimbrough, *supra* note 249, at 419.

258 Paul Hoffman, Ph.D., suggested this point. Moreover, the court in *Cavallo v. Star Enterprise* suggests a similar point. See 892 F. Supp. 756, 773-74 (E.D. Va. 1995). "[T]here may be instances where the temporal connection between exposure to a given chemical and subsequent injury is so compelling as to dispense with the need for reliance on standard methods of toxicology." *Id.* This comment suggests the court has an appropriately broad conception of pertinent evidence for toxic tort suits.

259 Such considerations make a plausible case for the appropriate causal claims and perhaps provide the best explanation among those available. Whether such an explanation is sufficiently good or certain enough to support the causation claim needed to convict a defendant for murder with toxicological evidence is more problematic.

260 One must not only show that the two deceased victims and the three sick victims had in common that they had ingested the lemonade (not necessarily an easy point to prove), but must also rule out other causes and show that the ingested substance was toxic. The Center for Disease Control was able to do this. See Kimbrough, *supra* note 249, at 418-19.

261 See, e.g., *Wade-Greaux v. Whitehall Lab., Inc.*, 874 F. Supp. 1441, 1475 (D.V.I. 1994); *In re Agent Orange Prod. Liab. Litig.*, 611 F. Supp. 1223, 1250 (E.D.N.Y. 1985).

262 See Wade-Greux, 874 F. Supp. at 1453.

263 See, e.g., Kimbrough, *supra* note 249 (describing cases where case studies on cancer in animals was useful in determining cause of death in humans).

264 See Cohrssen & Covello, *supra* note 236, at 27-48 (describing factors scientists examine in identifying hazardous chemicals).

265 See Anders Ahlbom & Maria Feychtig, *Studies of Electromagnetic Fields and Cancer: How Inconsistent?*, 27 Envtl. Sci. Tech. 1018, 1018-20 (1993); B. Hileman, *Findings Point to Complexity of Health Effects of Electric, Magnetic Fields*, 72 Chem. Eng. News 27, 33 (1994) (including related articles); D.A. Savitz, *Health Effects of Low-Frequency Electric and Magnetic Fields, Special Report Commentary*, 27 Envtl. Sci. Tech. 52, 54 (1993).

266 Janet Raloff, *Physicists Offer Reassurances on EMF: Electromagnetic Fields and their Link to Cancer Might be Tenuous*, 147 Sci. News 308, 308 (1995).

267 See *Health Effects of Low-Frequency Electric and Magnetic Fields: Special Report*, 27 Envtl. Sci. Tech. 42, 50-51 (1993); Raloff, *supra* note 266, at 308.

268 Compare Christopher Joyce, *Public Being 'Misled' Over Asbestos Dangers: Science Magazine Publishes Paper Claiming Public Health Risk*, *New Scientist*, April 14, 1990, at 16, 16, with B.T. Mossman et al., *Asbestos: Scientific Developments and Implications for Public Policy*, 247 Sci. 294, 298-99 (1990).

269 See Joseph Palca, *Lead Researcher Confronts Accusers in Public Hearing*, 256 Sci. 437, 437-38 (1992).

270 Compare James A. Bond et al., *Epidemiological and Mechanistic Data Suggest that 1,3-butadiene Will Not Be Carcinogenic to Humans at Exposures Likely to Be Encountered in the Environment or Workplace*, 16 Carcinogenesis 165, 165-71 (1995), with Ronald L. Melnick & Michael C. Kohn, *Mechanistic Data Indicate that 1,3-butadiene Is a Human Carcinogen*, 16 Carcinogenesis 157, 157-63 (1995).

271 For example, several scientists disagree on which substances constitute major human carcinogens. See Bruce N. Ames, *What Are The Major Carcinogens in the Etiology of Human Cancer?: Environmental Pollution, Natural Carcinogens, and Causes of Human Cancer: Six Errors*, in *Important Advances in Oncology* 237 (Vincent T. DeVita, Jr., M.D. et al. eds., 1989); Jean Marx, *Animal Carcinogen Testing Challenged*, 250 Sci. 743 (1990); Frederica P. Perera et al., *What Are the Major Carcinogens in the Etiology of Human Cancer?: Industrial Carcinogens*, in *Important Advances in Oncology*, *supra*, at 249.

272 See, e.g., Perera et al., *supra* note 271, at 252-59; H.J. Eysenck, *Were We Really Wrong?*, 133 Am. J. Epidemiology 429, 432 (1991) (arguing that multivariate analysis is essential to understand the relationship between smoking and cancer or coronary heart disease).

273 Sander Greenland, *Invited Commentary: Science Versus Public Health Action: Those Who Were Wrong Are Still Wrong*, 133 Am. J. Epidemiology 435, 435 (1991).

274 See Eysenck, *supra* note 272, at 429-32.

275 See Greenland, *supra* note 273, at 435.

276 See generally Phillip Kitcher, *The Advancement of Science: Science Without Legend, Objectivity Without Illusions* (1993). Sometimes science proceeds by building blocks, but at other times it seems to advance in paradigm shifts. See Amicus Curiae Brief of Physicians, Scientists, and Historian of Science filed in Support of Petitioners at 3-7, *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993) (No. 92-102).

277 See Cranor, *supra* note 104, at 29-48.

278 See *id.* at 29-48; see also *supra* notes 144-48 and accompanying text.

279 See generally Green, *supra* note 94; Sanders, *supra* note 74.

280 See *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 590-95 (1993).

281 Of course, our recommendation may result in some errors (false positives). However, in having a preponderance of the evidence burden of proof, the civil law system implicitly accepts the notion that some errors will be made because it requires nothing like certainty. And, as we have indicated two different kinds of mistakes can be made, legal false positives and legal false negatives, both of which the tort law should seek equally to prevent. See *supra* notes 144-48.

282 Plaintiffs providing a scintilla of evidence sufficient to survive an admissibility review, however, may not have evidence sufficient to obtain a directed verdict, even in the absence of any defense evidence.

283 For a detailed description of a sufficiency review in this context, see Sanders, *supra* note 74, at 1429-35.

284 See *supra* Part III.B.4.d.

285 See generally Green, *supra* note 94; Sanders, *supra* note 74.

286 See Huff, *supra* note 229, at 201.

287 See Filov et al., *supra* note 231, at 18-21 (discussing Krakovsky interspecies study which demonstrated “a linear correlation between the toxicity parameters and body weight” for eighty to eighty-five percent of the substances studied).

288 See *id.* Often these differences are widely publicized by special interest groups giving the public the mistaken impression that the results seen in experimental animals have little value for predicting adverse effects in humans. See *id.*

289 See Chan & Hayes, *supra* note 235, at 206-12; Mosberg & Hayes, *supra* note 235, at 226-31.

290 See *supra* note 236 and accompanying text.

291 See Huff, *supra* note 229, at 201; International Agency for Research on Cancer, Ethylene Oxide, in 60 IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 73, 139 (1994).

292 See Castleman, Regulations Affecting Use of Carcinogens, in Cancer Causing Chemicals 78 (Sax ed., 1981); David E. Lilienfeld, The Silence: The Asbestos Industry and Early Occupational Cancer Research--A Case Study, 81 Am. J. Pub. Health 791, 791-798 (1991); David Michaels, Waiting For the Body Count: Corporate Decision Making and Bladder Cancer in the U.S. Dye Industry, 2 Med. Anthro. Q. 215, 217-27 (1988); cf. Donald R. Mattison & John E. Craighead, Reproductive System, in Pathology of Environmental and Occupational Disease 559-72 (John E. Craighead ed., 1995).

293 The sedative and hypnotic drug thalidomide was used throughout the world to facilitate sleep and to reduce nausea and vomiting during pregnancy. The use of this drug by women in Europe and Asia during pregnancy resulted in the births of thousands of severely deformed children. See Jeanne M. Manson & L. David Wise, Teratogens, in Casarett and Doull's Toxicology: The Basic Science of Poisons 226, 227 (Mary O. Amdur et al. eds., 4th ed., 1991); James L. Schardein, Chemically Induced Birth Defects 228-38 (1993). Affected children typically exhibited missing limbs or limbs in which the long bones were dramatically shortened. Although this is occasionally cited as a failure of animal testing, see Chengelis & Gad, *supra* note 233, at 2, no testing of this drug for developmental toxicity was performed prior to its release into the European marketplace. See Schardein, *supra*, at 238. As stated by Chengelis and Gad: “Current testing procedures (or even those at the time in the United States, where the drug was never approved for human use) would have identified the hazard and prevented this tragedy”. Chengelis & Gad, *supra* note 233, at 3.

294 1,2-Dibromo-3-chloropropane (DBCP) was widely used in the United States as a soil fumigant and nematocide. In the mid-1970s, the wives of workers on a company softball team noted that many of the couples were having trouble conceiving. Clinical investigations revealed severely decreased sperm counts in the husbands which was subsequently shown to be caused by occupational exposure to DBCP. See Anthony R. Scialli & Michael J. Zinaman, Introduction to Reproductive Toxicology and Infertility xii, xiii (Anthony R. Scialli & Michael J. Zinaman eds., 1993). Similar effects were seen in others occupationally exposed to DBCP. See P.J. Gehring et al., Solvents, Fumigants, and Related Compounds, in *Handbook of Pesticide Toxicology* 637 (Wayland J. Hayes, Jr. & Edward R. Laws, Jr. eds., 1991). These effects eventually led to the banning of this pesticide in the United States. Toxicological testing conducted in the early 1960s had shown that this agent produced testicular atrophy in all species tested. See Mattison & Craighead, *supra* note 292, at 562. For reasons that are not clear, this

information was not used to protect workers exposed to DBCP, nor was a thorough evaluation conducted of the adverse reproductive effects of this agent on humans. See *id.*

295 Asbestos is a broad term used to describe a group of naturally occurring fibrous mineral compounds. Due to their resistance to thermal and chemical degradation, asbestos fibers have been widely used for insulation as well as other products including textiles, paints, plastics, roofing shingles, gaskets, brake linings, vinyl tile, and cement pipes. See Ernest Hodgson et al., *Dictionary of Toxicology* 42 (1988). Occupational exposures to asbestos particularly in mines and in the building trades have resulted in thousands of cases of lung cancer and asbestosis, a fibrotic lung disease. See Lilienfeld, *supra* note 292, at 791-92. Cancers of the pleura, peritoneum, bronchi and oropharynx have also been associated with exposure to this agent. See Hodgson et al., *supra*, at 42. Recent evidence has surfaced which indicates that asbestos was shown to cause lung cancer in laboratory animals approximately twelve years before the first human epidemiological evidence was published. See Huff, *supra* note 229, at 207; Lilienfeld, *supra* note 292, at 791-97. This bioassay information was suppressed until recently by the asbestos industry. See Lilienfeld, *supra* note 292, at 791-97.

296 For example, rats administered high doses of sodium saccharin in the diet exhibited increased frequencies of bladder cancer. Similar effects have not, however, been seen in humans consuming this artificial sweetener, although it should be noted that the human doses were considerably lower. Tumors induced in the kidney of male Fischer 344 rats by unleaded gasoline, 1,4-dichlorobenzene, and D-limonene developed due to an excessive accumulation of a specific protein, alpha 2 microglobulin, and resulting cytotoxicity in the rat kidney. See Gordon C. Hard et al., *Hazard Evaluation of Chemicals That Cause Accumulation of Alpha 2u-globulin, Hyaline Droplet Nephropathy, and Tubule Neoplasia in the Kidneys of Male Rats*, 99 Envtl. Health Persp. 313, 316 (1993). Tumorigenic effects are not seen in female Fischer rats or other tested animal species which do not exhibit the accumulation of alpha 2 microglobulin. Since humans appear to lack this specific protein and so not exhibit an accumulation of similar proteins in the kidney, the results observed in male rats are not considered to be relevant to humans. See *id.*

297 See Cohrrsen & Covello, *supra* note 236, at 49-50; International Agency for Research on Cancer, *supra* note 229, at 17-19.

298 See International Agency for Research on Cancer, *supra* note 229, at 11-12, for a more detailed description of this process for carcinogens.

299 See National Toxicology Program, *supra* note 44, at 3.

300 See Cohrrsen and Covello, *supra* note 236, at 49-50. The National Toxicology Program has a similar but not quite identical classification of carcinogens and bases its classification on "IARC's classification scheme and degrees of evidence." National Toxicology Program, *Seventh Annual Report on Carcinogens: Summary 1994* 6 (1994). The Annual Report's 'Reasonably Anticipated' category does not distinguish whether the degree of evidence supporting a given listing corresponds to the IARC categories of either 'Probable' or 'Possible' carcinogens. The text entries for listed substances, however, make clear whether the degree of evidence supporting the listing corresponds to either the 'Probable' or to the 'Possible' IARC category.

Id.

301 See International Agency for Research on Cancer, *supra* note 291, at 139.

302 See *id.*

303 See International Agency for Research on Cancer, *supra* note 229, at 17.

304 *Id.* (emphasis added). They add that "[t]he possibility that a given agent may cause cancer through a species-specific mechanism which does not operate in humans should also be taken into account." *Id.*

305 See *supra* notes 298-302 and accompanying text.

306 The idea is to take certain administrative or scientific bodies as providing rebuttably presumptive authoritative evidence on causation.

307 *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 780 (3d Cir. 1994). The court went on to note that animal studies satisfy three of the four Daubert factors: testability, following a generally accepted methodology, and being peer-reviewed in addition to

being used for purposes outside litigation. See *id.* at 781. The court concluded by holding “that the animal studies pass Daubert muster, are admissible and are one source by which plaintiffs can prove the harmful effects of PCBs.” *Id.*

308 See International Agency for Research on Cancer, *supra* note 229, at 17.

309 *Id.* at 17.

310 *Id.*

311 Since both plausibility and risk are problematic ideas, the weight of the combination would have to be assessed.

312 See Ward et al., *supra* note 173, at 539.

313 See *id.* at 542.

314 *Id.*

315 See *id.* at 539.

316 See *id.* at 547.

317 National Inst. of Envtl. Health Sciences, U.S. Dep't of Health and Human Servs., Seventh Annual Report on Carcinogens: Summary 1994 246-48 (1994).

318 Green, *supra* note 94, at 680-681.

319 See *supra* notes 244-47 and accompanying text.

320 See International Agency for Research on Cancer, *supra* note 229, at 18; Office of Research & Dev., U.S. EPA, Proposed Guidelines for Carcinogen Risk Assessment 85 (1996); U.S. EPA, EPA/625/3-91/019F, Alpha 2M-Gobulin: Association with Chemically-induced Renal Toxicity and Neoplasia in the Male Rat (1991); see also Hard et al., *supra* note 296, at 316.

16 VAELJ 1

End of Document

© 2019 Thomson Reuters. No claim to original U.S. Government Works.